

STIC-Biotech/ChemLib

60225

From: Rao, Manjunath N.
Sent: Wednesday, February 13, 2002 7:37 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/749,972

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10C 01
Phone: 306-5681

Date: 2-13-02

Please search the following as soon as possible for application with serial number **09/749,972**

SEQ ID NO: 1 against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO: 2, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
U. S. Patents and Trademark Office
Art Unit 1652, 10A11
Phone: 703-306-5681

Edward Hart
Technical Info Specialist
S. C. 12014
Conf 12014 Tel: 303-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 2/13/02
Date Completed: 2/15/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 1
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 04
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 22:13:42 ; Search time 1339.35 Seconds
(without alignments)
7870.756 Million cell updates/sec

Title: US-09-749-972-1

Perfect score: 639

Sequence: 1 agtcaaaaggatttttagt.....gaatggcttgccaaatga 639

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pl:**

25: em_ro:**

26: em_sts:**

27: em_sy:**

28: em_un:**

29: em_vi:**

30: em_htgo_hum:**

31: em_htgo_inv:**

32: em_htgo_rod:**

33: em_htg_hum:**

34: em_htg_inv:**

35: em_htg_rod:**

36: em_htg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	635.8	99.5	12815	1	AE007398
2	635.8	99.5	111135	2	SPNEU1906
3	253.8	39.7	10183	1	AE005502
4	246.6	38.6	12560	1	AE006276
c 5	227.4	35.6	3466	1	AF178425
6	187.2	29.3	300150	1	AP001507
7	166.6	26.1	180136	1	BAC180K
8	166.6	26.1	213080	1	BSUB0001
9	131	20.5	301050	1	AP003130
10	131	20.5	343590	1	AP003359
11	109.6	17.2	971	1	MC109
12	107.2	16.8	49617	6	AX067453
c 13	103.2	16.2	10471	1	AE004722
c 14	89.6	14.0	12333	1	AE001292
c 15	86.2	13.5	16207	1	AE002314
16	83.2	13.0	311321	1	NMA322491
c 17	82.8	13.0	347660	1	AP002994
18	81.6	12.8	10075	1	AE002421
19	81.6	12.8	349980	6	AX044030
c 20	77.6	12.1	11763	1	AE008073
21	77.6	12.1	13037	1	AE001769
c 22	76.4	12.0	40544	1	SCH5
23	73.4	11.5	13594	1	AE006204
c 24	72.6	11.4	14530	1	AE004276
25	71.4	11.2	327650	1	MPULM01
26	71	11.1	14567	1	AE000716
c 27	70.8	11.1	9932	1	AE001612
28	70.8	11.1	11176	1	AE002210
29	68.2	10.7	633	6	AR139594
c 30	68.2	10.7	10238	1	U32728
c 31	66.4	10.4	329100	1	SME591787
32	64.6	10.1	11010	1	AE001873
c 33	62.2	9.7	2658	1	LLARCAB
34	56.6	8.9	12059	1	U39679
c 35	55.6	8.7	11308	1	AE002101
c 36	55.4	8.7	495	1	YEN270424
37	54.2	8.5	495	1	YEN270427
c 38	52.8	8.3	279110	1	RFX03
39	52.6	8.2	495	1	YEN270429
40	52.4	8.2	1694	1	AF065312
41	52.2	8.2	495	1	YEN270423
42	52.2	8.2	495	1	YEN270425
43	52.2	8.2	6109	1	AF294440
44	50.6	7.9	495	1	YEN270426
45	49	7.7	495	1	YEN270428

ALIGNMENTS

RESULT 1
AE007398
LOCUS
DEFINITION Streptococcus pneumoniae section 81 of 194 of the complete genome.
ACCESSION AE007398 AE005672
VERSION AE007398.1 GI:14972397
KEYWORDS
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 12815)
Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,
McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and


```
/transl_table=11
/product="glutamate 5-kinase"
/protein_id="AAK75055.1"
/db_xref="GI:14972406"
/translation="MKYKRIVFKVGTSSLTNEGSLRSKVKDITQOLAMLEHAGHEL
ILVSSGAIAGCAIGFKRPKTIADKOASAAVGGGLLEEVYTNLLRLQIVSAQILL
TODDFVDRKRYRNAQALSVLNRGAIPRIINENDSVVIDELKVGNDLUSQAAMVQ
ADLLFTLDVGLTGNPNSDPRAKRLRITINREIIDMAGAGSSNGTGMGLTKIK
AATIAATESGPIYICSLKSDSMIEAAEETEDGSYFAQEKGRITQKWLAFYAQSOG
SIWVKGAALSOYKSLLSGIVAEAGVFSYGDIVTFVDRKSKSLKGRVQFGAS
ALEDLRSOKAKGVLIYRDDWISITPEIQLLTFE"
gene
7739..9001
/genes="SP0932"
7739..9001
/genes="SP0932"
/codon_start=1
/transl_table=11
/product="gamma-glutamyl phosphate reductase"
/protein_id="AAK75056.1"
/db_xref="GI:14972407"
/translation="MYSRQEQFEQVQVKKISINTASEEVKNQALLAMADHLVAATEEI
LAANLDMAAKGKISDVMLDRLYLADRIEAMRGIREVVALPDP IGVLETSOLEN
GVITTKRVAMGVIGIIVESRPNVTSDAALTLKSNVAVLSRGDAYQTTTHAIVTAL
KGLLETTIHPNVLIVEDTSRESSYAMKAKGYLDLLIPRGAGLINAVVENALVPV
IETGTVIHVVYDKDAEDKALSIINNAKTSRPSVCNAMEVILLHENKAASFLPRLEQ
VLVAERKEAGLEPIQIRLDSKASQFVGOAETQDFTEFLDVLAVKYVSSLEAVA
HIESHTHSDAIVTENAEAAAYFTDOVDSAAVYNASTRFTDGGQFGLGCEMGISTQ
KLHARGPMLKELTSYKYVVVAGDQIRE"
gene
9005..9802
/genes="SP0933"
9005..9802
/genes="SP0933"
/notes="identified by match to PFAM protein family HMM
PF01089"
/codon_start=1
/transl_table=11
```

```
Query Match          99.5%; Score 635.8; DB 1; Length 12815;
Best Local Similarity 99.7%; Pred. No. 1.8e-168;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcaaaaggatttttagtctctcttggaggaccagaggagcagcagaccagtgtt 60
Db 9918 ATGTCAAAAGGATTTTGTCTCTCTGTGAGGGACCAAGAGGAGCAGCAAGACAGTGT 9977

QY 61 tttagagctctgctaccatttttagaggaagagtagagtgatttgacagaccgtgaa 120
Db 9978 TTAGAGGCTCTGTACCAATTTTAGAGGAAAAGGAGTAGAGTGTGAGCACCCTGAA 10037

QY 121 cctggcggagctctgattggggagagattcgggaagtagatttttgatcccaagtcatact 180
Db 10038 CCTGGCGGAGCTTTGATTGGGGAGAGATTTCGGGAAGTAGATTGGATCCAAAGTCATACT 10097

QY 181 cagatggatgctaaacagagctactctctctatatcttcagtcgcagacagcatttggtg 240
Db 10098 CAGATGGATGCTAAAAACAGAGCTACTTCTCTATATTGCCAGTCGCAGACAGCATTTGGTG 10157

QY 241 gaaaagtcttcacacccttgagctggcaagtttgggtcatcagatgattttatcgat 300
Db 10158 GAAAAGAGTCTTCCAGCCCTTGAAGCTGGCAAGTTGGTCATCATGTGATCGTTTATTCGAT 10217

QY 301 agttctgttgccctatcaggagatttggctggtcttagatattgaagccattgactggctc 360
Db 10218 AGTTCYGTGGCTATCAGGGATTTGGTCGTGGCTTAGATATTGAAGCCATTGACTGGCTC 10277

QY 361 aatcagtttgcgacagatgggctcaaacccgatttgacactctattttgacatcgagggtg 420
Db 10278 AATCAGTTTGCACACAGATGGCCCTCAAAACCGATTTCACACTCTATTTTGACATCGAGGTG 10337

QY 421 gaagaaggctggctgattgctgcgaatagtgaccgcgagagtttaactcgtttggatttg 480
Db 10338 GAAGAAGGGCTGGCTGCTATGTGCTAATAGTAGTGACCGGAGGTTAATCGTTGGATTG 10397
```

```
QY 481 gaagggttgagactgcataaaaaagttctgcaaggctacacattctctcttgataaagag 540
Db 10398 GAAGGGTTGACTTGCATAAAAAAGTTCTGCAAGGCTACCTTCTCTCTTCTGGATAAAGAG 10457

QY 541 ggaatgcgattgtcaagattgatgtctctcccttggagcaagtttggaactacc 600
Db 10458 GGAATGCGCATTTGCAAGATTGATGCTAGTCTCCCTTTGGAGCAAGTTGTGGAAACTACC 10517

QY 601 aaggctgtctgtttgaacgaatgggcttggccaaatga 639
Db 10518 AAGCTGTCTCTTGTGACGAATGGGCTTGCCCAATGA 10556

RESULT 2
SPNEU1906 SPNEU1906 111135 bp DNA HTG 11-JUL-2001
LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
DEFINITION in ordered pieces.
ACCESSION AL449928
VERSION AL449928.1 GI:11545153
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae.
REFERENCE 1 (bases 1 to 111135)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francisco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae
JOURNAL type 19F clinical isolate
MEDLINE Microb. Drug Resist. 7 (2), 99-125 (2001)
REFERENCE 21355329
AUTHORS 2 (bases 1 to 111135)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francisco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..111135
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 32875 a 20454 c 25072 g 32680 t 54 others
ORIGIN
```

Query Match 99.5%; Score 635.8; DB 2; Length 111135;
Best Local Similarity 99.7%; Pred. No. 2.5e-168;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 atgtcaaaaggatttttagtctcttggaggaccagaggagcagcagaccagtgtt 60
Db 73721 ATGTCAAAAGGATTTTGTCTCTCTTGGAGGACCAAGAGGAGCAGCAAGACAGTGT 73780

QY 61 tttagagctctgctaccatttttagaggaagagtagagtttgacagaccgtgaa 120
Db 73781 TTAGAGGCTCTGTACCAATTTTAGAGGAAAAGGAGTAGAGTGTGAGCACCCTGAA 73840

QY 121 cctggcggagctctgattggggagagattcgggaagtagatttttgatcccaagtcatact 180
Db 73841 CCTGGCGGAGCTCTTGTGTTGGGAGAGATTTCGGGAGTAGATTTCGGATCCAAGTCATACT 73900

QY 181 cagatggatgctaaacagagctactctctatatcttcagtcgcagacagcatttggtg 240
```

||||| 73901 CAGATGGATGCTAAACAGAGCTACTTCTATATATGCGATCGACACATTTGGTG 73960
Qy 241 gaaaaagtctctcagccctgaagctggcgaagtgtgcatcatggtatcgcttttcatcgat 300
||||| 73961 GAAAAGTCTTCAGCCCTTGAGCTGGCAAGTTGGTCATCATGGATCGTTTATCGAT 74020
Qy 301 agttctgtgctatcagggtattgctgctggtgcttagatattgaagccattgaactgctc 360
Db 74021 AGTTCTGTGCTATCAGGAGTTTGGTCGTGGCTTAGATATTGAAGCCATTGACTGGCTC 74080
Qy 361 aatcagittggacagatgggtctaaacccgatttgacactctattttgacatcgaagtgtg 420
||||| 74081 AATCAGTTTGGCAGATGGCTCTCAAAACCCGATTGACACCTATTGACATCGAGGTG 74140
Qy 421 gaagaaggctggtctgattgctgcgaatagtgacgcgaggttaatcgtttgattgtt 480
Db 74141 GAAGAAGGCTGGCTGTATGCTGCTAATAGTACCGCGAGGTTAATCGTTTGGATTG 74200
Qy 481 gaagggttgactgcatataaaaggctcgtcaaggctacattctctctggtataaagag 540
Db 74201 GAAGGTTGGACTTGCATAAAAAATTCGTCAAGGCTACCTTCTCTTCTGGATAAAGAG 74260
Qy 541 ggaatcgcattgtcaagattgatctagctccctctggtgagcaagttgtggaactacc 600
Db 74261 GGAATCGCATGTGCAAGATTGATGCTAGTCTCCCTTTGGCAAGTGTGGAACATACC 74320
Qy 601 aaggctgctgttgcggaatggcttgcccaaatga 639
Db 74321 AAGGCTGCTTGTGTTGACGAATGGCTTGCCCAAAATGA 74359

RESULT 3
AE006502
LOCUS AE006502 10183 bp DNA BCT 01-JUN-2001
DEFINITION Streptococcus pyogenes M1 GAS strain SP370, section 31 of 167 of
the complete genome.
ACCESSION AE006502 AE004092
VERSION AE006502.1 GI:13621639
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Streptococcus pyogenes M1 GAS.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10183)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, O., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE 11296296
PUBMED
REFERENCE 2 (bases 1 to 10183)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, O., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA
FEATURES
source 1..10183
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
115..744
/gene="upp"
/note="SPy0392"
115..744
/gene="upp"

/EC_number="2.4.2.9"
/note="Best Blastp hit = sp|P36399|UPP_STRSL URACIL
PHOSPHORIBOSYLTRANSFERASE (UMP PYROPHOSPHORYLASE)
(UPRTASE) >gi|347998|gb|AAA26890.1| (L07793) uracil
phosphoribosyltransferase [Streptococcus salivarius]"
/codon_start=1
/transl_table=11
/product="putative uracil phosphoribosyltransferase"
/protein_id="AAK33430.1"
/db_xref="GI:13621640"
/translation="MGKCOVISHPHQKLSILRROTTSTKDFRELNVETAMLMGVYEV
SRDLPLEVDVDTQTPVSKYQKOLAGKLAIVPILRAGICGMVDGILLSVPAKVGCHGM
YRNEETLEPEVYLKLPEDINQRIELVDPLATGSAIILAVDSLAKRGRANIKFVCL
VAAPGEGVKLQEAHPDIDIFTAALDHLNHEHGVIVPGLDAGDRLFGTK"
713..847
/gene="SPy0393"
713..847
/gene="SPy0393"
/note="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33431.1"
/db_xref="GI:13621641"
/translation="MLVIVYLQNKRDNHQIVAFGLGQSFLILFKKSERFLFFKGKASL
"
968..1558
/gene="clpp"
/note="SPy0395"
968..1558
/gene="clpp"
/EC_number="3.4.21.92"
/note="Best Blastp hit = sp|P36398|CLPP_STRSL
ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT
(ENDOPEPTIDASE CLP) >gi|347999|gb|AAA26891.1| (L07793)
ATP-dependent protease proteolytic subunit [Streptococcus
salivarius]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent protease proteolytic
subunit"
/protein_id="AAK33432.1"
/db_xref="GI:13621642"
/translation="MIPVVIETQTSRGSYDIYSRLIKDRILMLTGPVEDNMANSVIA
QLLFLDAQNTKDIYLYVNTPGGSVAGLAIVTMFIRADVOTIVMGMAASMGTVIA
SSGTCGRKRLPNAEYKIHOPGSGTGGTQOTDMAIAAEHLKTRHRLKILAQNAK
TKQIHKDAERYWMSAEETLAYGFIDEIMENNELK"
2050..2325
/gene="SPy0397"
2050..2325
/gene="SPy0397"
/note="Best Blastp hit = pir||D69874 conserved
hypothetical protein ylbG - Bacillus subtilis
>gi|2340003|emb|CAB1353.1| (298682) ylbG protein
[Bacillus subtilis] >gi|2633871|emb|CAB13373.1| (299111)
similar to hypothetical proteins [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33433.1"
/db_xref="GI:13621643"
/translation="MFQKQERIGLVVLYNRRDARKLSKFGDLYHSHKRSRYLIYIN
KNDLTKLEMRRLKCVKDIRPSAFDDIDRQFVGNLHRDETNNHKG"
2574..3209
/gene="tmk"
/note="SPy0399"
2574..3209
/gene="tmk"
/note="Best Blastp hit = dbj|BAB03761.1| (AP001507)
thymidylate kinase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative thymidylate kinase"

```
/protein_id="AAK33434.1"
/db_xref="GI:13621644"
/translation="MITGKLITVEGPDGAGKTTVLQELPLLKQKVAQDILLTTREPGG
VAISEHRELILINDIAMDPKTELLLYTAAROHLEVEKVLPALEAGOLVIFDRFIDS
SVAYQAGRLIKADIQWLNEKATDLEPLDTLYFDVPSIGLARINAOQRENRDL
LEFIEIHORVKCYLALAKEHPKRIIVTIDATKPLKEVSVALEHVLALLA"
3227..4102
/gene="SPY0400"
CDS
3227..4102
/gene="SPY0400"
/notes="Best Blastp hit = dbj|BAB03763.1| (AP001507) DNA
polymerase III delta' subunit [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative DNA polymerase III delta' subunit"
/protein_id="AAK33435.1"
/db_xref="GI:13621645"
/translation="MDLAOKAPNVYQAFOTILKKDLNHNAYLFSGDFANEMALFLAK
VFCEKQKQTPGCHRSQOLIEQGFADVTVEPTGQVIRKTDVVREMANFVSQTGTE
NRQVFILIKDCDKMHINANSLLKYIEEPQGEAYIEFLTNDNDNKVLPTIKSRQVQFQ
PKNEAYLYLQAOEKGLINHOAKIVAKLATNTSLERLQTSKLELITQAEFVSIVL
KQDQAYLALNRLVQALATEKEQDLVLTLLTLARERAQTLTQLEAVYQARLMQWS
NVNFQNTLEYVMNSE"
4121..4360
/gene="SPY0401"
CDS
4121..4360
/gene="SPY0401"
/notes="Best Blastp hit = dbj|BAB03764.1| (AP001507) signal
peptidase-like protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33436.1"
/db_xref="GI:13621646"
/translation="MGRKSLSKDLPVATKVAHKKKAPENTAFATDLMKEYSRLIT
QNLQLEMTFVNSVFLLDQHVVLIFPSAEHRMDFNC"
4765..5088
/gene="SPY0405"
CDS
4765..5088
/gene="SPY0405"
/notes="Best Blastp hit = dbj|BAB03765.1| (AP001507)
unknown conserved protein in B. subtilis [Bacillus
halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33437.1"
/db_xref="GI:13621647"
/translation="MNKLELFDGFSQNLMTVLAIEAMKKOVOSLSVEENTILRLE
NTKLRLERLSLHEHTVAKNPSKQRKHLEGIYDEGFHICNFFYQORRENDEECMFCE
LLDRK"
5093..5956
/gene="SPY0406"
CDS
5093..5956
/gene="SPY0406"
/notes="Best Blastp hit = dbj|BAB03768.1| (AP001507)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33438.1"
/db_xref="GI:13621648"
/translation="MOVOKSFDKTKTSGTLVLPPTIGNLQDMTERAVATLKVEWDFIC
AEDFRNTGLLKHFDIATKOISFEHNAYEKIPOLDLLISGRSLAQVSDAGPMSISD
PGHDLVKAISDSIAVVALPGASGITALIAGLAPQPHVYGLFPRKAGQQAFFED
KHHPYQMPYESPYRKDTLTNNLACYGDQVVLVRLFKLEFYQSGISRLISYL
EETPLKGECLLIVAGAQAQSEVELTADVDLVSLVQKEIQAGAPNOAIKTIAKAYQVN
RQELYQQFHD"
5983..6375
/gene="SPY0407"
CDS
5983..6375
/gene="SPY0407"
/notes="Best Blastp hit = emb|CAA61744.1| (X89566)
```

```
cytochrome-c oxidase sbunit I [Nitrobacter winogradskyi]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33439.1"
/db_xref="GI:13621649"
/translation="MTEDQATILILGVYIILAVFLSALLCLIANRRLPKRAGYKGM
YSLIPFYAGFVKHKITFGEDNKNFYFIFGVIATIIYVYVHFCYVRAFGSKGLAVFGLF
FPGIATSLIVAFGKTYTTQECHPVPHLLN"
complement(6422..7051)
/gene="SPY0408"
CDS
complement(6422..7051)
/gene="SPY0408"
Query Match 39.7%; Score 253.8; DB 1; Length 10183;
Best Local Similarity 65.4%; Pred. No. 1.le-60;
Matches 388; Conservative 0; Mismatches 202; Indels 3; Gaps 1;
Qy 1 atgtcaaaagatttttagtctctctgagggaccagagggagcagcgaacaccagtgtt 60
Db 2574 ATGATCACAGGAAACTCATAAACCGTTGAAGTCCAGATCGCGACGGGAAACGACTGTT 2633
Qy 61 ttgagggtctgtctaccaatttttagagggaaaaagagtagagg---tgttgaccacctt 117
Db 2634 TTAGAGCAACTCATTCCTACTGTTAAACAAAAAGTGGCTCAGGATATCTTGACAAACA 2693
Qy 118 gaacctggcaggtcttgattggggagagaagattcgggaagtgtatttgatcccaagtc 177
Db 2694 GAGCTTGAGGTTGGCCATTTCAGAACACATTAAGAGAACTATATTTAGACATCAACCAT 2753
Qy 178 actcagatggatgctaaacacagctactctctatatattgccagtcgcagcagcagcttg 237
Db 2754 ACTGCTATGGACCCCTAAACAGAACTCTTACTTTATATTTGGCTCGAAGGCAGCATTTG 2813
Qy 238 gtgaaaaagttcttccagccctgaagctggcaagttgggtcatcatggtatctttatc 297
Db 2814 GTGAAAAGGTACTACCAGCTCTTGAAGCAGCCAGTTGGTTTATTGATCGTTTCATC 2873
Qy 298 gatgttctgtgctatcagggtattgtgtcgtggttagatattgaagcactgactgg 357
Db 2874 GATAGTTCAGTAGCTTATCAAGGTCGCGGACGCGGTTTGTGATCAAGAGACATTCAGTGG 2933
Qy 358 ctcaatcagttgcagcagctggtctcaaacccgattgacactctatttggacatcgag 417
Db 2934 CTCACGAGTTTCGCAACAGATGTTTAGAGCCTGACTTGACCCCTTATTTGATGTTCTT 2993
Qy 418 gtgaaagaagggtcggtcgctattgtcgcgaatagtgaccgcgaggttaactcgtttgat 477
Db 2994 TCTCAAAATTGTTTAGCAGCGATTAAACGCGAACCAACAAAGAGAAGTCAATCGTTTAGAT 3053
Qy 478 ttgaaagggttggtgacttcataaaaaagttcgtcaaggtcacccttctctctgataaa 537
Db 3054 TTGAAAACAATAGAGATACATCAACGAGTAGCAAAAGGATACCTTAGCCCTTAGTAAAGAG 3113
Qy 538 gagggaaacgattgtcaagattgactgtctctctctcttggagcaagtgtt 590
Db 3114 CATCCAAAACGATTGTGCACAAATTGATCGGACTTAACACCTTAAGAAAGATTGT 3166
RESULT 4
AE006276
LOCUS
DEFINITION Lactococcus lactis subsp. lactis IL1403 section 38 of 218 of the
complete genome.
ACCESSION AE006276 AE005176
VERSION AE006276.1 GI:12723260
KEYWORDS
SOURCE Lactococcus lactis subsp. lactis.
ORGANISM Lactococcus lactis subsp. lactis
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
```

REFERENCE 1 (bases 1 to 12560)
 AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
 Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
 TITLE The Complete Genome Sequence of the Lactic Acid Bacterium
 Lactococcus lactis ssp. lactis IL1403
 JOURNAL Genome Res. 11 (5), 731-753 (2001)
 MEDLINE 21235186
 PUBMED 11337471

REFERENCE 2 (bases 1 to 12560)
 AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
 Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de
 Villvert, Jouy en Josas 78352, France
 FEATURES
 source
 1..12560
 /organism="Lactococcus lactis subsp. lactis"
 /strain="IL1403"
 /sub_species="lactis"
 /db_xref="taxon:1360"
 129..320
 /gene="ydljB"
 /note="L195067"
 129..320
 /gene="ydljB"
 /note="EVIDENCE BY GENEMARK BIO16.01 UNKNOWN. General.
 homologs not detected in Streptococci."
 /codon_start=1
 /transl_table=11
 /product="UNKNOWN PROTEIN"
 /protein_id="AAK04486.1"
 /db_xref="GI:12723261"
 /translation="MIHELTDRAHQHFSTRKIDAFWIFEPTTCALFEGGLVSENKAI
 TFEINAKTLEVFENEKIL"
 362..679
 /gene="trxB"
 /note="L195348"
 362..679
 /gene="trxB"
 /note="EVIDENCE BY HOMOLOGY BIO02.10 Thioredoxin,
 glutaredoxin, and glutathione. similar to the thioredoxin
 H-type from plants."
 /codon_start=1
 /transl_table=11
 /product="thioredoxin H-type"
 /protein_id="AAK04487.1"
 /db_xref="GI:12723262"
 /translation="MIIPENIENLAELVKGPEKTVFFFTAGWCDCNFIKPKMPEIEV
 ENPDFKEVDREIDMLAIEWGIPSFVIEDGKEKARLVNKLKRTKEEVNTFLA
 SAK"
 765..1391
 /gene="ydljP"
 /note="L195751"
 765..1391
 /gene="ydljD"
 /note="EVIDENCE BY HOMOLOGY BIO15.01 HYPOTHETICAL. homolog
 to YtpR B. subtilis, has local similarity to similarity to
 phenylalanine tRNA ligase"
 /codon_start=1
 /transl_table=11
 /product="HYPOTHETICAL PROTEIN"
 /protein_id="AAK04488.1"
 /db_xref="GI:12723263"
 /translation="MIATYNTNVGDVLMILITANDKGEKVNFRKGVKVARVFIETGAT
 VANNIFEAKSLLANSGQVVLSSDIETLNKELSGSDVLVHDAQKFFVVAEIV
 EMEDHDSHLHCKVNVGFADPQIVCGAPNAALGKLTVAALPGAMPGNSLIIFPGA
 LRGVQSYGMLCSARELPNAPEVRGIIELSSDLSAGKAFDSETMWOA"
 1554..2894
 /gene="noxE"
 /note="L196579"
 1554..2894
 /gene="noxE"
 /note="EVIDENCE BY HOMOLOGY BIO06.01 ENERGY METABOLISM.

Aerobic"
 /codon_start=1
 /transl_table=11
 /product="NADH oxidase"
 /protein_id="AAK04489.1"
 /db_xref="GI:12723264"
 /translation="MKIVVIGTNGHAGIATANTLIDRYPGHEIVMIDRNSNMSYLGCGT
 AIWVGRIEKPDELFYKAEDFERKGVKILFETEVSEIDFTNKKIYAKSKTGEKITES
 YKLVLATGSRPIIPNLPKDLKGIHFLKFOEQOADEEFAKNDVKRIAVIAGIYIG
 TEIAEAAKRRGKVELLPDAGKSLSYVDEFEAKGMENLAHQHTELFHGLAQEFKA
 NEKHVSQIVTNKSTYDVLVINCIGTANSALAGEHLETFKNAIKVKDKHQQSSDPD
 VSAVDATIIYSNAKQDFTYIATALASNAVRSGIVAGHNHIGGKSIESVGOVNGSIFG
 YNMTSTGLSKAAKIKGLEVSFDFEDKQAWFLHNNDSVKIIRIVETKNRRIIGAQ
 LASKSEIIAGNIMFSLAIQEKKTIDELALLDLFLPHFNSPYNTMYTVALNAK"
 2984..3373
 /gene="ssbA"
 /note="L0299"
 2984..3373
 /gene="ssbA"
 /note="EVIDENCE BY HOMOLOGY BIO10.02 DNA replication,
 restriction, modification, recombination, and repair. THIS
 PROTEIN IS INVOLVED IN DNA RECOMBINATION AND REPAIR"
 /codon_start=1
 /transl_table=11
 /product="single-strand binding protein"
 /protein_id="AAK04490.1"
 /db_xref="GI:12723265"
 /translation="MNKTMILIGRLTSAPEISKTNDKSYVVTAVNRFRKNEKGERE
 ADFSIIWKSATETLSYAKKGLSISIEGIRNTYDKONQKHVYTEILGLSLYDLL
 ESRATIALRESAITEETLLDAEELP"
 3493..3777
 /gene="groES"
 /note="L198515"
 3493..3777
 /gene="groES"
 /note="EVIDENCE EXPERIMENTAL PMID:8486277 BIO04.03
 Chaperones"
 /codon_start=1
 /transl_table=11
 /product="10 KD chaperonin"
 /protein_id="AAK04491.1"
 /db_xref="GI:12723266"
 /translation="MLKPLENVVLVRKVEEESKSMGGIVILTSASQEKPTAEVAVGE
 GKTNNHGTLSPLVKVGDVIFEFKAGTIVKMDGEELILKDSLLAIVE"
 3865..5493
 /gene="groEL"
 /note="L198893"
 3865..5493
 /gene="groEL"
 /note="EVIDENCE EXPERIMENTAL PMID:8486277 BIO04.03
 Chaperones"
 /codon_start=1
 /transl_table=11
 /product="60 KD chaperonin"
 /protein_id="AAK04492.1"
 /db_xref="GI:12723267"
 /translation="MSKDIKFSDDARTAMRGIDILADTVKTTLTPGKRVNVLEKSYG
 SPLITNDGVTIAKEIELEDFHFNKAGLVSEVASKTNDIAGDGTITATVLTQAIYREG
 LKNVTAGNPVIRGIRGELAAETASIKEMAPVHDKSAIAQATVYSSRSEKVEYII
 SDAMERVSGDGVITIEESGMQTELDVVEGQFDGRYLSQTMVSNTEKMWAELENDPYI
 LITDKKISNIEQLPLEQLIKTNRPILLIIVADDVDGEALPTLVLNKIKGVNVAVKA
 PGFGDRRAQOELDIALITGGTVITEELGDLKDATLEALGOAAKATVDKDHITTEVEGA
 GSDAI SDRAVITKAOIKETKTSDFDREKLOERLAKLAGVAVKVGAAATTELEKAMKL
 LIEDALNATRAAEEGIVSGGCTALVNAIALDKLSEGDIOGTGINVRRALEPVRQ
 IANAGTEGVIIDKLRSEVGVTGNAATGQVNMIEEGIVDPKAVTRSALQNAASVA
 GLIITTAEVANKPEPAAPAMPMPDPSMGGM"
 complement(5539..6351)
 /gene="yeaA"
 /note="L582"
 complement(5539..6351)
 /gene="yeaA"
 /note="EVIDENCE BY HOMOLOGY BIO15.01 HYPOTHETICAL.
 homology to yycJ B. subtilis"

		/codon_start=1 /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="AAK04493.1" /db_xref="GI:12723268" /translation="MTEQNGFKYSILASGTGNSFYLETPEPKKILYDAGLSGKKIESL LSEIDRIDHIDALLTHESHDHKIGVIARKYIGDLYANEDTWSELDSKNAKGIP VEQKHUFEMKMTGFDIDIESFGVSHDAIAPQFYFMKDNKSNFVMLTDTGYVSDMKR GTENADYIELMNSHDIELRGAQYWPRTKRIQLSDGHLSDNEADGAFAAFSIHGNTKK RIPLGHLSLNNIKELAHLTMDGYLQQHDVDTKKDLKILDTISHEASQLYDI" /complement(6522..7964) /gene="kinC" /note="L0124" complement(6522..7964) /gene="kinC" /note="EVIDENCE EXPERIMENTAL PMID:9172368 BIO09.02 Two-component systems" /codon_start=1 /transl_table=11 /product="sensor protein kinase" /protein_id="AAK04494.1" /db_xref="GI:12723269" /translation="MIKFFNSMIFKGVAHLILFVLFMSSEFDYNAHLSTFTGYGLF SLVFLVFLNQOHMIVNVYSKOMNDIKNGYLNDRDKPKGTSEINELYDELKIGEDIV RTQNLENORNOLDLSLAYVMQGVATDRRGNTIMANKSALHYLNTDNLIMOKNIVE VLKTDQYWFYDLLEKEPTIETHDSGNEISLHKFAIFRSGESFGIITAVLHDM TEQDKAERENRFLVSNVSEHLRFTPLSVRAYLEADGAIEDKETASSFINVSLTFDM RMIMISDLTUSLSDMPDQKRVILNKEMINLIATFDYQIINRLDQILETESNDLISSTFTL VRYISQPIWEIETDKQIAQVIDINIGNAIKYSPKGGTITISVEKRESVDLISISQDQ MGTPKSDLLKIFDRYRVONASRNSKVGGTGLGLSLIVHDIKIHGCTIFASSEGEKGK TFTSFETPLYSLEAADDDMDQDEFEDNE" complement(7957..8658) /gene="llrC" /note="L0125" complement(7957..8658) /gene="llrC" /note="EVIDENCE EXPERIMENTAL PMID:10708382 BIO09.02 Two-component systems. involved in acid stress resistance development." /codon_start=1 /transl_table=11 /product="two-component system regulator" /protein_id="AAK04495.1" /db_xref="GI:12723270"
Query Match	38.6%	Score 246.6; DB 1; Length 12560;
Best Local Similarity	63.0%	Pred. No. 1.2e-58;
Matches 381; Conservative	0; Mismatches 224; Indels	0; Gaps 0;
QY	7	aagagatttttagtctctgtgaggaagagtagtaggtgttgacgacccgtgaacctggc 126
Db	8839	AATGGAATCTAAATTCCTTTGAGGACCGCGATGGAGCAGGAAAAACAACGGTTTTTAAA 8998
QY	67	gctctgcatacaatttttagaggaaagagtagtaggtgttgacgacccgtgaacctggc 126
Db	8899	GAATATTTTACCTGAATCCAAAAAATAAGCGCGAAATTCGCCAATCTGTAACCAAGC 8958
QY	127	ggagctctgattggggagaagattcgggaagtgatttttgatccaagtcatactcagatg 186
Db	8959	GGTGTTTCGAGTGGCGGAAGAAATTCGCCAAATTAATCTTGATCCAAAGAAATACAGATAT 9018
QY	187	gatgctaaacaagctactctctatattgccagtcgcagacagcattgggtggaaaaa 246
Db	9019	GATTCATAAACCCAGGTTAATGCTTTTTCGCGCGCGCGTCTGTTTACATATATGCAAGAAAA 9078
QY	247	gttcttcacgaccttgaagctggcaagttggtgcatactgatgctgttttctcgatgtct 306
Db	9079	ATGTTGCGCAGCTTTACAAGCGGGTAAGATGTGTCATTGTTGACCGATTATTGATTCACT 9138
QY	307	gttgcctatcagggatttgcgtgccttagatattgaagccattgactggctcaatcag 366
Db	9139	GTCCCTTATCAGGGGTATGCCGTGATPTTGGGTCTCGAAGTTGTGATTGGCTTAACAT 9198

Qy	367	tttgcacagatgggtccaaacccgattgacacactctattttgacatcgaggtggaagaa	426
Db	9199	TTTGCACAGATGGTGTGAAGCCCTGATTGACGCTTTATTTGATGTTGATACGGATGTT	9258
Qy	427	ggcgtggtcgtattgtgtggaatagtagacccgcgaggttaacgttttgagattgagaagg	486
Db	9259	GCCCTTGAGCGAATTTGAAAAATTCGAGCTGATGAGGTAAATCTGTTAGATTTCGGAAGA	9318
Qy	487	ttggacttcataaaaaagttcgtcaaggctaccttctctctctggtgataaaggaggaat	546
Db	9319	GCTGAGATGATCGTAAGTTCGTGAAGTTATTTCGGAGATTGTTGTAAGAAGACCTTCAG	9378
Qy	547	cgcatttcgaagattgatctagctctccctcttgagcaagttgttgaaactaccaaagct	606
Db	9379	CGTTTTCGAAAAATAGATGCAAGTCAACCTTTAGAAAAAGTGTAGCTGATACACTTCC	9438
Qy	607	gtctt 611	
Db	9439	GTCTT 9443	
RESULT	5		
AF178425/c			
LOCUS	AF178425	3466 bp	DNA BCT 25-JAN-2001
DEFINITION	Lactococcus lactis TcsCorf1 (tcsCorf1984), Rrc (rrc), and Kinc (kinc) genes, complete cds; and TcsCorf2 (tcsCorf2) gene, partial cds.		
ACCESSION	AF178425		
VERSION	AF178425.1	GI:12483715	
KEYWORDS	Lactococcus lactis.		
SOURCE	Lactococcus lactis.		
ORGANISM	Bacteria; Firmicutes; Bacillus/clostridium group; Streptococcaceae; Lactococcus.		
REFERENCE	1 (bases 1 to 3466)		
AUTHORS	O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G.F., Ehrlich, S.D. and Morel, P.		
TITLE	Six putative two-component regulatory systems isolated from Lactococcus lactis subsp. cremoris MG1363		
JOURNAL	Microbiology 146 (Pt 4), 935-947 (2000)		
MEDLINE	20244638		
PUBMED	10784052		
REFERENCE	2 (bases 1 to 3466)		
AUTHORS	O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G.F., Ehrlich, S.D. and Morel, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-AUG-1999) Department of Microbiology, National University of Ireland, Cork, Western Road, Cork, Ireland		
FEATURES	Location/Qualifiers		
source	1..3466		
	/organism="Lactococcus lactis"		
	/strain="MG1363"		
	/db_xref="taxon:1358"		
gene	complement(171..806)		
	/gene="tcsCorf1984"		
CDS	complement(171..806)		
	/gene="tcsCorf1984"		
	/note="thymidylate kinase-like"		
	/codon_start=1		
	/transl_table=11		
	/product="TcsCorf1"		
	/protein_id="AAG53731.1"		
	/db_xref="GI:12483716"		
	/translation="MNGILSLGPGGAGKTVVLQELPEIQMKREVVPTREPGGVRVAEEIQIILDPKNETIDSKETLMLFAARLHQMKLPALRAGKVIIVDRFIDSSVAYGYQRIIDGVEVDVNFATDGLKPLDTLYFDITDVALERIMKRADEVNRLLDRAEMHRKREGYLEIVAKEPGREVKIDASQSLKVVADTILEVLKRFVSEF"		
gene	984..1685		
	/gene="rrc"		
CDS	984..1685		
	/gene="rrc"		
	/note="putative response regulator"		
	/codon_start=1		

TITLE Sequencing of three lambda clones from the genome of alkaliphilic
 Bacillus sp. strain C-125
 JOURNAL Extremophiles 3 (1), 29-34 (1999)
 MEDLINE 99184646
 REFERENCE 8 (sites)
 AUTHORS Takami,H. and Horikoshi,K.
 TITLE Analysis of the genome of an alkaliphilic Bacillus strain from an
 industrial point of view
 JOURNAL Extremophiles 4 (2), 99-108 (2000)
 MEDLINE 20263314
 REFERENCE 9 (sites)
 AUTHORS Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,
 Hirana,C., Fuji,F. and Takami,H.
 TITLE Characterization and comparative study of the rrn operons of
 alkaliphilic Bacillus halodurans C-125
 JOURNAL Extremophiles 4 (4), 209-214 (2000)
 MEDLINE 20426005
 REFERENCE 10 (sites)
 AUTHORS Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,
 Fuji,F., Hirana,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and
 Horikoshi,K.
 TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis
 Nucleic Acids Res. 28 (21), 4317-4331 (2000)
 JOURNAL 20512582
 MEDLINE 11 (bases 1 to 300150)
 REFERENCE Takami,H. and Takaki,Y.
 TITLE Direct Submission
 JOURNAL Technology Center, Deep-sea Microorganisms Research Group; 2-15
 Submitted (22-MAR-2000) Hideto Takami. Japan Marine Science and
 Technology Center, Deep-sea Microorganisms Research Group; 2-15
 Natsushima, Yokosuka, Kanagawa 237-0061, Japan
 (E-mail:takami@jamstec.go.jp,
 URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,
 Tel:81-468-67-3895, Fax:81-468-66-6364)
 FEATURES Location/Qualifiers
 source 1..300150
 /organism="Bacillus halodurans"
 /db_xref="taxon:86665"
 /note="alkaliphile"
 584..1933
 /gene="dnaA"
 584..1933
 /gene="dnaA"
 /note="BH0001"
 /codon_start=1
 /transl_table=11
 /product="initiation of chromosome replication"
 /protein_id="BAB03720.1"
 /db_xref="GI:10172613"
 /translation="MENIIDLWELRALKMEKKVSKPSFETLWTKQTKANSTEDSTIIIT
 APNEFARDWLEKHDELISEITDIDLTVGLRYPKVPVPTSQLDPEPVEQLKPKMKOPP
 AONGENPNMLNDKYTFDTFVIGSGNRFHAASLAVAEAPAKAYNPLFYGVGLGKT
 HLMHAIGHVMDHPNPNKVVYLSSEKFTNEFINAIRDKNKAVNRKNYRVNVDVLLIDDI
 QFLAGKEQEEFFHTFNALHEDNKQIVISSORPKPEITPLEDRLSRFEWGLTIDIT
 PPDLFTRIALIRKKAENLDIPNVTANOIDNIREGALIRVAYASLNLND
 MNADLAELKIDIPNAPKPRVLTITDIQKTVGVEFHVKLEDFKAKRTKSVAPPROIA
 MYSRLSLTASLPKIGSEFGGRDHTTVIHAHEKISKLLSTDQELQKIDISDKLRS"
 2103..3245
 /gene="dnaN"
 2103..3245
 /gene="dnaN"
 /note="BH0002"
 /codon_start=1
 /transl_table=11
 /product="DNA polymerase III (beta subunit)"
 /protein_id="BAB03721.1"
 /db_xref="GI:10172614"
 /translation="MHFVIDRDFVQNVHNSKAVSSRTTIPILIGIKIVADHEGVTLL
 TGSDSISITETPLPEGRQNVKEGSIVLQAKFAEIVKKLPEQETETIHVDQSF
 VTIRSGSSVFNGLDPDEYRLPLVLEHDHVRLPQKILKDIIRQTVFVASTQETRP
 VLTGVNFIEDGLITCTATDSHRLANRKYVPVKNDDELQFSNVNIPGKSLNELSKILD
 ENEELDDIVVTENQTLFKLKNMLFFSRLEGKYPVTKNMPKEAKTSFAVHTKAFLOT
 LERALLSREGNQVINLKTLDGVVVEVTAITPEIGKVTENVATQGLEGEELRISFNG

gene
 CDS
 3429..3650
 /gene="BH0003"
 3429..3650
 /gene="BH0003"
 /note="BH0003"
 unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB03722.1"
 /db_xref="GI:10172615"
 /translation="MSQQLPITTSYITLQQLLKEVGVIDTGGMAKWLAEEHVLNVE
 SEDRRGKLFSGDSVTIADGTVIDLVDAAE"
 3662..4777
 /gene="recF"
 3662..4777
 /gene="recF"
 /note="BH0004"
 /codon_start=1
 /transl_table=11
 /product="DNA repair and genetic recombination"
 /protein_id="BAB03723.1"
 /db_xref="GI:10172616"
 /translation="MHIERUTLKQFNVDLYSFPFNVIIGENAGQKTNVIEAIY
 FLAKASHRTARDKELTQWEAPFARIEGAFQKQNGPLSLHVLGSGKKGKVKVNGLEQR
 RLSDYICAVNVVMFGPDLNLVRGSPQIRRRFLDMELGQSPVYLHQLAMYQILQQR
 NLLKOLFGRKANDPMLDVLTDOLIEVAVETKRRFEITQLQORWAEIHOAITSRGKE
 KLVTYEPSVHVSEOLNLSKLREGFYQAEYOKKERSEQRGTTLGGHDDDLVRFVNDK
 DVQTVSGQQRRTALSKLAELMEKTEVGDYPILLDDVLSELDYRSHLLHLAIO
 HRVQTVTTNVGDIDHQTLOEATYIEVEQGSUKKKG"
 4788..5057
 /gene="BH0005"
 4788..5057
 /gene="BH0005"
 /note="BH0005"
 unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB03724.1"
 /db_xref="GI:10172617"
 /translation="MFIHLGGDHVIRSKDVIAILDRDMEQSSGVTKYLEHHOKNGNI
 VTISDLIKSIVVTPDTYIYSPISSTYTKRRSHAVSELDRFGEE"
 5130..7043
 /gene="gyrB"
 5130..7043
 /gene="gyrB"
 /EC_number="5.99.1.3"
 /note="BH0006"
 /codon_start=1
 /transl_table=11
 /product="DNA gyrase subunit B"
 /protein_id="BAB03725.1"
 /db_xref="GI:10172618"
 /translation="MTREQQAYDESOIQVLEGLAEVVRKPGMYIGTSSRGLHLVWE
 IVDNSDEAMAGFCDGPIVEEDNSITVDNGRIPGVGHEKMGPAVEIVMTVLHA
 GKGFGGGYKVGSLHGVASVYNALSTMLEVHEHREGKVHYQKFKRGVPAADLEVG
 TTDRTGKTHFKPDGDIETFTTVFYDVLASRLAFLNKLRIKKTDMREEEKSDE
 FHYEGGIASEVHEHLNRTRETLHETPIHIEGKQGVVEIAVQYNDGFTSNLYSFANNI
 NTHGGTHESGFKGLTRVINDYARKHNLPKESDNLGTEDREGTLAIIISVKIPDPQ
 FEQGTKILGNSPARTISLTFEHSRFLIEHPQVARKLVLDKGLMASRAREAKKAR
 ELTRRKSALENSLPGLKADSSRDASISRIYVEGDSAGSAGQGRDRHFQAILPLR
 KTLNVKARLDKILANNEIRAITALTGTGDDFDIEKARYHKIITMDADVDGAHI
 RTLLTFEYMRPLIEHGVYIAQPLKYVQGGKIOYAGDAEQELIISQLPDKAK
 PGIQRYKGLGEMNPSQLWETMDPGQGTVLQVLTEDAMKADEIFEILMDRVEPRDRF
 IQENARYVNLDI"
 7065..9566
 /gene="gyrA"
 7065..9566
 /gene="gyrA"
 /EC_number="5.99.1.3"
 /note="BH0007"
 /codon_start=1
 /transl_table=11

```
/product="DNA gyrase subunit A"
/protein_id="BAB03726.1"
/db_xref="GI:10172619"

Query Match          29.3%   Score 187.2:   DB 1;   Length 300150;
Best Local Similarity 57.3%;   Pred. No. 1e-41;
Matches 339; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

Oy 1 atgtcaaaagatttttagtctctttagggagaccagggagcagcagcaaacagcagttt 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60731 ATGACGAAAGCTGTTTATTACAGTTGAAGTGGAGAAGCGCAGCAAAACGCTCTGCC 60790

Oy 61 ttagagggctcgtaccatttttagagaaagagtagaggtgttgacgacccgtgaa 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60791 CTAGATCGGATTAGGAAATCTTAAGAGAGAACGCCCTATCGCTGACGACCGCTGAG 60850

Oy 121 cctggcagctctgattgggagagattcggaaagtgttggattccgaagtcatact 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60851 CAGGGGGAATCCCATTTGCTGAGCAAAATTCGTCATCATATTTGGATGTTGACCATACA 60910

Oy 181 cagatggatgctaaacagagctactctctctatattgcagtcgcagacagcatttggtg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60911 AGAATGGACCGAGACAGAGCATTTGCTTTATGCAGCAGCCGTAGACACGACCTGGTT 60970

Oy 241 gaaaagtcttcacgacctgaagctggcgaagttgttcacatcgattgttttcatcgat 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60971 GAAAAGTGTTCAGCAGCACTGAAGCAGGCGATGTCGTCCTATGCGATCGCTTCATTGAT 61030

Oy 301 agttctgtgctatcaggagatttggtgtgctgattagattgaagcattgaactggtc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61031 AGTAGTTGGCGTATCAGGAGATATGCAAGAGGTATTTGGGTTTGAGGATATTTAGCGGATA 61090

Oy 361 aatcagttgctgcagatggctcaaaccccgatttgacactctatttggacatcgaggtg 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61091 AATGAATTTGCCATTAAGGACGCTATTCGACACTTAACCTTTTCGTTGCTGATGCCA 61150

Oy 421 gaagaaggctggtctgattgctgcgaatagtgaccgcgaggttaactcgtttggattg 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61151 GACGTAGGCTTATCAGCAATTCATCGTGATCAGTCAAGAGACAAATCGCTTGACCAA 61210

Oy 481 gaagggttgacttgcatataaaagttcgtcaagctacaccttctctcttgataaagag 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61211 GAGGCCCTTACGTTTCACCAAAAGTAAAGAAGGTTATGAACCTATATGTCGACACGTAC 61270

Oy 541 ggaatcgctgctcaagattgctagctctcctcttcttggagcagttgtgtg 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61271 CCCAGACGAGTAGTGGAGATTGATGCACCAACACAGTCCCTTTGATCAAGTTGTTG 61322

RESULT 7
BAC180K      BAC180K      180136 bp      DNA      BCT      06-FEB-1999
LOCUS       B. subtilis DNA, 180 kilobase region of replication origin.
DEFINITION  D26185
ACCESSION   D26185
VERSION     D26185.1  GI:467326
KEYWORDS    .
SOURCE      .
ORGANISM    Bacillus subtilis (sub_species:Marburg, strain:168) DNA.
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
            1 (sites)
            Moriya,S., Ogasawara,N. and Yoshikawa,H.
            Structure and function of the region of the replication origin of
            the Bacillus subtilis chromosome. III. Nucleotide sequence of some
            10,000 base pairs in the origin region
            Nucleic Acids Res. 13 (7), 2251-2265 (1985)
            2 (sites)
            Alonso,J.C., Shirahige,K. and Ogasawara,N.
            Molecular cloning, genetic characterization and DNA sequence
            analysis of the recW region of Bacillus subtilis
            Nucleic Acids Res. 18 (23), 6771-6777 (1990)
            91088245

REFERENCE
AUTHORS     Ogasawara,N. and Yoshikawa,H.
TITLE       Genes and their organization in the replication origin region of
            the bacterial chromosome
JOURNAL     Mol. Microbiol. 6 (5), 629-634 (1992)
MEDLINE     92204018
REFERENCE   4 (bases 1 to 180136)
AUTHORS     Ogasawara,N., Nakai,S. and Yoshikawa,H.
TITLE       Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
JOURNAL     DNA Res. 1 (1), 1-14 (1994)
MEDLINE     96051385
REFERENCE   5 (bases 1 to 180136)
AUTHORS     Ogasawara,N.
TITLE       Direct Submission
JOURNAL     Submitted (24-DEC-1993) to the DDBJ/EMBL/GenBank databases. Naotake
            Ogasawara, Nara Institute of Science and Technology, Graduate
            School of Biological Sciences; 8916-5 Takayama-cho, Ikoma, Nara
            630-01, Japan [E-mail:nodasawa@bs.aist-nara.ac.jp,
            Tel:07437-2-5430, Fax:07437-2-5439]
FEATURES
            Location/Qualifiers
            1..180136
               /organism="Bacillus subtilis"
               /strain="168"
               /sub_species="Marburg"
               /db_xref="taxon:1423"
               complement(1691..2398)
               /codon_start=1
               /transl_table=11
               /product="unknown"
               /protein_id="BAA05173.1"
               /db_xref="GI:467327"
               /translation="MDKKILVDDKPIADILEFNLRKEGYEVHCAHDGNEAVEMVEE
               LQPDLLIDIMLPKNGQVEYCRVKYDMPILMTAKDSEIDKVICLGADDYVTK
               PFSTRELLARVKANLRQLTAPAEESPSSNEIHGSLVIFPDAYVYVSKRDEITLTH
               REFELHLAKHIGQVNTREHLLQTVMGYDFGDVTDVTVRRLREKIEDNPSPHPNM
               IVTRRGYYILRNPEQD"
               complement(2782..2854)
               /product="tRNA-phe"
               complement(2890..2966)
               /product="tRNA-asg"
               complement(3048..3119)
               /product="tRNA-asg"
               complement(3129..3204)
               /product="tRNA-Lys"
               complement(3429..4721)
               /gene="purA"
               complement(3429..4721)
               /gene="purA"
               /codon_start=1
               /transl_table=11
               /product="adenylosuccinate synthetase"
               /protein_id="BAA05174.1"
               /db_xref="GI:467328"
               /translation="MSSVVVGTQWGDGEGKKITDFLSENAEVIARYOGGNNAGHTIK
               FDGTYKHLIPSGIFVKDKTCVIGNMVDPKALVTELAYLHERNVSTONLRISNRA
               HVLLPYHLKLDVEERKGGANKIGTTKGGIPAYMDKAAIRIGIRIADLLDRDPAEKL
               ERNLEKNRLLKEMYTEGFKLEIDILDEYYEYEQQIKKYVCDTSVVLNDALDEGRRVL
               FEQAQGMLDIDOGTTPFTVSSNPAGGVGTIGSGVGTPIKHKHVVGVSGITLSDISIDL
               FPELKDIDGQIREVGREYGTTPRRRVGVDFSDVVVHARVSGTTLKSLNSIDL
               AGTELKICVARYKGEIIEFPASLKALAECEPVEEMPQGTEDITGAKSLSELSEN
               ARHVLERSOLTGIPLSIFSVGPDRSQTNVLRVYRAN"
               complement(4927..5346)
               /codon_start=1
               /transl_table=11
               /product="unknown"
               /protein_id="BAA05175.1"
               /db_xref="GI:467329"
               /translation="MGKRFSFOAAQIARIAPTQGLDEIIRFYEEGICLKRIGEFSOH
               NGDGYWFGCLPHADYHLEFQYEGGSTAPVPHDPSLLVFPVNAVELAATSLKIHMG
               IQEVESENPYWSMGVGTIEDPDGWRIVFMNSKGISGK"
               complement(5467..6831)
               /gene="dnaC"
```


CDS

```
complement(5467..5831)
/gene="dnaC"
/codon_start=1
/transl_table=11
/product="replicative DNA helicase"
/protein_id="BAA05176.1"
/db_xref="GI:467330"
```

```
/translation="MTDLLNDRLPPONIEBQAVIGAIFLPQSALTALASEVILPDDFY
RMSHKIYNAMVLGDRGPEVDLVTVSELANTDLLEEVGISYLTIDIANSTPAANI
EYAKIIEKSLRLRTATTIAQGYTREDEVDLLSEAEKTIWEVAQRKNTSAFQ
NIKDLVOTYDNEIOLYNRKGDITGPTGFTELDRMTAGFORNDLIIVAARPSVGTAK
FALNIAQNVATKTESVALEFSLEMGAEOLVMEMLCAEGNINAQNLRTGNLTEDRGKL
TWAMGSLNSGIYIDDTGPIRSEIRAKCRRLKQBSGLGMILIDLQLIQGSGRDN
RQGEVSEISREKSTARELQVPIALSQLSRGVEQQRKPMMSDIREGSIIEQADAI
VAFLYRDYDKETENKNIIEIIIAKQRNPVGTVSFLAFVKEFKFVNLERFDDAGV
PGA"
```

CDS

```
7001..7201
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05177.1"
/db_xref="GI:467331"
```

```
/translation="MIKKYAITPNVDADGWFEIVENVAFTALYTSKDAAIERAKQVAK
ENSPSKLVIYDQFKNVEEHSF"
```

CDS

```
7447..7713
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05178.1"
/db_xref="GI:467332"
```

```
/translation="MHKNPPYLFYNGHPPIALFLCDIFLHFTSDCTMDRREGEFVR
PQIQAETAQKLAESLNDLPLEQIMHPQHILLAKWAELOKEDKS"
```

CDS

```
7786..8994
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05179.1"
/db_xref="GI:467333"
```

```
/translation="MPHPHNKKIQSFWLITGIIIFIAFNLRPAITSVGPVIISIRAEHL
MNGAAGFTLPLLSFAVLSPKLGORLGNERTLWGLVILLIGVLTGSTGYTAA
LFGFALIGVTAIGNVLPSLKHKYPEKGMISLSTYTSNMNIPAAIASGYSVPLAT
OMNGWKQAFLLWGLLALLIWIIPOLBHRDTANQTMKLOSSSIWASKMANYVIFM
GLQSLFYSLIAWFFELLKSHGIDTATAGMWSLMFOFSLPSTFLPVLADRVKQGR
IVAAALSYLIGLGLAGGSHLLAIWMIIIGIGGSSISLALILIGLSRANAQQA
ALSGMSQFGYLAAVGPFIYGVLFEDQTHSWTMPVILLIAALIVMGAAGQAGRDYI
FOSEKORNSA"
```

CDS

```
9099..11156
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05180.1"
/db_xref="GI:467334"
```

```
/translation="MKTREDAALILILAAAFNLNTHWQDDTANOVYLAAVKSWTO
SFHNFYASDFPSGFYVDKPPVWLVIQIFALIFGVHTWSVIIQALAGVSGVELLY
RWKPTFGVGAARIALNVALPIAVASRTNNIDSMVLFTLLGLSTCLLRVAKGKL
WMLLTAFALIGAFKAMNMMAFWLPFAFVLYLIASRSVKKKKIGSLLSLVLTGLS
LSWAJAVDSASSSPRYGSSQTSNVLFAFGYNGTFLRGQTTGLAKGDMNAAGGN
MONQDMAPNGSFSFONGNOSHQAOPPPNGOTGALNGGGTPTTGGNGPGN
GGPGGGGKSVNMFGTGDPLRLFQSALSQOLSMMLPESLIGLIGALIIWSYDRGRH
AKKMETLEWALTPVAGFFSIAGFHOYYLIMLAPPIALSGIGWTMHRLYRNKK
DNSSYLLPAVALITAVFQYIILSAVTSQIGSVMMYVLGLLIGITLALLMKRSHPS
KLLTISICVLVLLTPYVSATPLLYXGNSVLPESGQLKSGSANGNMFSSEVDSGLLS
YLRKNTGEYLFATLTPVTAAPYIIYENESVMAMGGFNGTDPILTVSELKKLIEKG
VYFLLSENNSGNSLVSWKKNKEITSDEYSGSSSTNSVQGMRRRGPAGESQOTLY
LVE"
```

gene

```
complement(11193..11642)
/gene="rplI"
complement(11193..11642)
/gene="rplI"
```

CDS

```
/codon_start=1
/transl_table=11
/product="ribosomal protein L9"
/protein_id="BAA05181.1"
```

CDS

```
/db_xref="GI:467335"
/translation="MKVIELQDVKGKKGKVENKVNADGYAHNFLIKKKGLAVEANASNI
SALNGQKQKEKEATAELQAKSLKETLKLTVLSEKSGEGRGLSGSVTSKQITQL
OKDHNIKVDKRLKLEPDGIRALGYTNVPVKLHPVQAVLKVVHKEEA"
complement(11639..13618)
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05182.1"
/db_xref="GI:467336"
```

CDS

```
/translation="MPSFYEKPLFRYPIYALIALSITILISFYFNWILGTVEVLLLA
VILFFIKRADSLIRQIDAYISTLSYRLKKVSEEALEMPIGIMLNDQYYIEWANPF
LSCSFNESTLVGRSLDTCESVYPLIQEVSETVTLNDRKFRVTKRDERLLYDFDV
TQIQIEKLYENERVTIAYIFLDNYDDVTQGLDDQRTSMNSQVTSLLNAAQOYEIF
LKRTSSERTIAVNLHAYITLFDNVDYQGLDDQRTSMNSQVTSLLNAAQOYEIF
LGDALQSLDLAGLRGGDOVAIKLPNGKVFYKGNPMKEKTRVRARVISHALKKIV
TESSNVIIMGHKFPDMDSIGAAIGLKVAQNNKDGFIVIDPNQIGSSVQRLIGEIKK
YEELWSRFTIPEAMEISNDNLLVIVDTHKPSLWMEERLYNKHIEHIVIDHRRGEE
FIRDPULVMEPIASSTAEVLTELEYQPKRUKINMIEATALLAGIIVDTKFSLRKG
SRTFDAASYLRAGADTVLVQKFLKETVDSYIKRAKLIQHTLVYKDNIAIASLPNEE
EYFDQVLIQAADSLMSSEVASFAVRDEQTVCSIRSLSGEVNVQIIMEALEGGG
HLTNAATOLSGISVSSEALERLKHAIIDEYFEGGVQR"
```

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

Qy 550 attgtcaagattgatgctagtagtctccctttggagcaagtgtg 590
 || ||||| | | ||| |||
 Db 103305 TTTCATTCAAGTTGATGCGGGCAATCCAAGACCTCGTCGT 103345

RESULT	8						
BSUB0001		BSUB0001	213080 bp	DNA	BCT	26-NOV-1987	
LOCUS		<i>Bacillus subtilis</i>	complete genome	(section 1 of 21):	from 1 to		
DEFINITION		213080.					
ACCESSION		Z99104	AL009126				
VERSION		Z99104.1	GI:2632267				

Bacillus subtilis.
Bacillus subtilis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus.
 1 (bases 1 to 213080)

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, L., Boursier, L., Brans, A., Braun, M., Brignelli, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devane, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Etian, K.J.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaier-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardonais, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Noaki, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetle, D., Porwollik, S., Prescott, A.M., Prescean, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T., Scanlan, E., Schleicher, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H., and Danchin, A.

TITLE
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)

JOURNAL MEDLINE
98044033

REFERENCE
2 (bases 1 to 213080)

AUTHORS
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission

TITLE
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

JOURNAL

```

FEATURES             Location/Qualifiers
     source            1..213080
                        /organism="Bacillus subtilis"
                        /strain="168"
     gene              /db_xref="taxon:1423"
                        410..1750
                        /gene="dnaA"
     CDS               /gene="dnaA"
                        410..1750
                        /function="Initiation of chromosome replication (DNA
                        synthesis)"
                        /note="alternate gene name: dnaH, dnaJ, dnaK"
                        /codon_start=1

```

```

/transl_table=11
/protein_id="CAB11777.1"
/db_xref="GI:2632268"
/db_xref="SWISS-PROT:P05648"
/translation="MENILDLNQAALQAEKKLSKSPFETWVKSTKAHSLQDGLTIT
APNEFARDWLESRYLHLIADIITVELTGEELSIFKIVPQMDQEDFMPKQPVKAKVED
TSDPQNLMPKPTFTDVISSGNRFAHAASLAVAEPKAYNPFLFIYGGVGLGTHFL
MHAITHGVVLDHNPSKAVVYLSSEKFFNEFINSRDNKAVDFNRNRYRNVDLVLLIDTTP
LAGRQTOEEFFHTFNLHEESKOIVISSDPKPEPTLEDRLRSRFEWGLITDITPP
LDREARIALBKKAKAEGLDIPNEVMLYIANIDSNIIELEGALIRVAVYSSLLINKDN
ADLAERALKDIISSRPKVITIKIQRVVQQQFNKLEDFKAKKRKTSKVAFFPQRTAMY
LSRMDTDSLLPKGEEPRGDRHTTVIHAHEKTSKLLADDEQLQQHVKEIKEQLK"
/length=3106
/gene="dnaN"
CDS
/feature="dnaN"
/EC_number="2.7.7.2"
/function="DNA synthesis"
/note="alternate gene name: dnaG, dnaK"
/codon_start=1
/transl_table=11
/product="DNA polymerase III (beta subunit)"
/protein_id="CAB11778.1"
/db_xref="GI:2632269"
/db_xref="SWISS-PROT:P05649"
/translation="MKFTIQKDLVESVODLVKAVSSRTTIPILTIKIVASDDGVSF
TGSDDSISEFTPKEGDKEIYVTIEQGSIVLQARFSEIVKKLPMAVTVEIVONQY
LTIIRSKAENLNGLDADBYPHLPQIEHHAIQIPTDLLKNLIQTVFAVSTSETRP
ILGVNIVSQSELCTATDASHRLAKRKLQIDPERSYNVSPKSLTELSKILDDND
QELVDIVITQVLFKAKNVLFPSRLLDNGYDPTTSLIPQDSKTEIIVNTRFLQAQID
RASLLAGERNNVVYKLSAKPAESIEISSNSPEIGKVVEAIVADQIEGELNLSFPKY
MLDALKVLEGAETRVSTGAMRFLRTNDETIVQILLPVPTY"
/length=3087
/terminator
/gene="dnaN"
/feature="dnaN"
/length=3206
/gene="yaaA"
CDS
/feature="yaaA"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB11779.1"
/db_xref="GI:2632270"
/db_xref="SWISS-PROT:P05650"
/translation="MANPISIDEMTITLQFLKLADVIQSGCMKWFUSEHVLVND
PDNRGRKLYGDVVEIEGFGSPQVYN"
/length=3437
/gene="recF"
CDS
/feature="recF"
/function="DNA repair and genetic recombination"
/codon_start=1
/transl_table=11
/protein_id="CAB11780.1"
/db_xref="GI:2632271"
/db_xref="SWISS-PROT:P05651"
/translation="MTIQNLSTSYRNDHAELOFENKVNVLIGENAOGKTNLMSEALY
VLMSKSHRTSDNKLRTMDKDYAKIEGRVMKQNGAIPMLQVLSKRGKRGKVNHLIEQ
KLSQYVGLNTIMFAPEDNLNVKSGPQVRRLDMEIGQVSPVYLDLSLYOKILSOR
NHFLLQQLQRTKQDTRMLDVLTDLQVEAAKVVVKRLQFTQALEKWAQPIHAGISRGL
EELTLKYHYALDVSDDLKSLGDSVQEAESKLREKEIERGVTLSGPHRDDVLFYVNG
RDVQTVCSQQQRTTALSCLKAEIDLIIHEEIGEYPIILLDDVLSELDDYRQSHLLHTI
QGRVQFTVITLSDVDGIDHETLROAGMFRVQNGALVK"
/length=4567
/gene="yaaB"
CDS
/feature="yaaB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB11781.1"
/db_xref="GI:2632272"

```

```
/translation="MYIHLGDDFVSTRDIVGLDFKANMSPIVEFLKKQKHVVPS
VNRPNLS"
gene      4866..6817
          /gene="gyrB"
CDS       4866..6782
          /gene="gyrB"
          /EC_number="5.99.1.3"
          /function="Initiation of replication cycle and DNA
          elongation"
          /note="alternate gene name: novA"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase (subunit B)"
          /protein_id="CAB11782.1"
          /db_xref="GI:2632273"
          /db_xref="SWISS-PROT:P05652"
          /translation="MFOQONSVDENQIQVLEGLAEVRRKPMGVIGSTNSKGLHLVWE
          IVDNSIDEALAGCYTDINTQIEKDNSITVVDNGRGPICVIEHKMKRPVAVETVTLVHA
          GKFDGSGYKGGVLGCVASVYNALSTELDVTRHGRGKIHQTKYRGVFDVLEIIG
          EYDHTGTTTFHVPDPFISSETTEYDYLANKRVRELAFUTKGVNITIEDRGSEGRKN
          EHYEGGTSKVEYVYLNRSKEVHEEPIYIEGEKDGITVEVALQYNDSTYSNIYSFTNN
          INTYEGTHAEAGFKTGLRVINDIYARKKGLIKENDPNLSCDDVREGTLAISIKHPDP
          QFEGQTKLGLSNLEARTITDLPSTAMETFMLENPDAAKKIVDKGLMAARAAWAKKA
          RELTRKSALEISNLPCKLADCSKDPSSSELYIVGEDSAGSAGKGRDRHFQAILPL
          RKILNVEARDKILSNNEVRSMITALGTGIGEDFNLEKARKHKVYIMTDADVDGAH
          IPTLLTFYRYMRQIIEGVIYIAOPPLYKVOQGRKVEYAYNDKELELLKTLPTTP
          KPLQRYKGLGEMNATQLWETTMPPSSRTLLQVLTLEDAMDADETFELMGLDKVPEPRN
          FTEANARYVKNLID"
          6792..6817
          /gene="gyrB"
          6993..9485
          /gene="gyrA"
          6993..9458
          /gene="gyrA"
          /EC_number="5.99.1.3"
          /function="Initiation of replication cycle and DNA
          elongation"
          /note="alternate gene name: nala"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase (subunit A)"
          /protein_id="CAB11783.1"
          /db_xref="GI:2632274"
          /db_xref="SWISS-PROT:P05653"
          /translation="MSEQNTPOVREINISQEMRTSLDYAMSIVYSRALPDVDRGLKP
          VRRIRIYAMNDLGMTSDKPYKKRSARIVGEVIGKYHPGDSAVVESVMRAQDFNRYM
          LVDHGNGFSGVDSAAAMRYTEARMSKISMETLRDITKDTIDYQNDYGSREPEVVM
          PSRFNLLNGRAGIAGVGNATNIPPHQLGEIDGVLVASENDPITIPELMEVPGDFV
          PTAGQLLGRSGINKAYESGRSITIRAKAEIETSSGKERIIVTELPYOVNNAKLIEK
          IADLVDRKKIEGIDLRDESRTGMVRIETIRRDANANVLNNLYKOTALQTSFGINL
          LALVDGPKVLTLKQCLEHYLDHQKVIIRRTAYELRKAERAHILLEGRLVALDHDLA
          VISLRNSQTAETARTGLTEQFSLTEKQAAILDMRLQRLTGLEREKIEEYQSLVKL
          IAEKDLIANEVKVLPIREELTEIKERENDEPTEIVTSGLETIEDEDLIERENLVV
          TLTHNGYVRLPASTYRSOKRGKGQVQMGNTDEDFVEHLISTSTHDTILFSSNKKV
          YRANGKEIPEYRTAGPIINLLEVEKGEWINAIIPTVEFNALYFFFTTKHGVSKR
          TSLSQFANIRNGIALISREDDLMVRLTGTQKIIITGKNGLLIRPEPTDVREMG
          RTAAGYKGLTLDVDDVVVGMEILEEESHVLIIVTEKYGKRTPAEYRTQSRGKGKLT
          AKITENNGOLVAVKATKGEEDLMIITASGVLIRMDINDISITGRVTVQGVRLIRMAEE
          HVAIVALEKNEEDENESEGEV"
          9465..9485
          /gene="gyrA"
          terminator
```

```
Query Match      26.1%; Score 166.6; DB 1; Length 213080;
Best Local Similarity 55.4%; Pred. No. 64e-36;
Matches 322; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY      10 gatttttagctctcttagggaccagggagcgaagcaccagtggttttagagcgt 69
         || || || || || || || || || || || || || || || || || || ||
Db      39163 GGTTTATTATTACATTCGAAGTCCTCGAAGTGCAGGAAACGACTGTCGAGGAG 39222

QY      70 ctgctaccaatttagagaaagagtagtagagtggttgacgacccqtagaacctgcgga 129
         || || || || || || || || || || || || || || || || || || ||
Db      39223 ATCAAAAACATACTGACAGAGAAGCGCTTCAGGTTATGCGAACTCGAGAGCCAGCGGA 39282
```

```
QY      130 gtcttgattggggagaagattcgggaagtgtatttttgatccaagtcatactcaagatgat 189
         || || || || || || || || || || || || || || || || || || ||
Db      39283 ATAGATATTCCAGAACAAATTAGAGAGGTTATTCTGTAATGAGAACAAATATATTATCGAC 39342

QY      190 gctaaacacagactactctctctatatattgcagtcgcgcagacagcagcatttgggtggaaaaagatt 249
         || || || || || || || || || || || || || || || || || || ||
Db      39343 CCAAAAACAGAAAGCTCTCTTTATATGCGGCTGCAAGACGCCAGCATTTAGTTGAAAAAGTA 39402

QY      250 ctccagcccttgaagctggcgaagtgtggtcatcatggttcttattcagatagtcagttctgtt 309
         || || || || || || || || || || || || || || || || || || ||
Db      39403 AACCTGCTTTAGAACAGGGGTTTATCGTTCTTTGCCACAGATTCATAGACAGTCCCTCTT 39462

QY      310 gctatcagggatttggctgctggttagatattgaaagccattgactggtcctcaatcagttt 369
         || || || || || || || || || || || || || || || || || || ||
Db      39463 GCGTATCAAGGATATGCAAGAGGGCTTGAATTCATGATGAGGTTCTGTCCATTAATGAATTT 39522

QY      370 gcacagagatgggtctcaaacccgatttgacactctatttgcacatcgagggtggagaagagg 429
         || || || || || || || || || || || || || || || || || || ||
Db      39523 GCCATTGGAGATATGATGTCGCCCATGTAAACGGTTTATTTTCAATTGATCCGGAAGAAGGA 39582

QY      430 ctggctctgattgctgcgaatagtgaccgcgaggttaactgcttggagtttgggaagggttg 489
         || || || || || || || || || || || || || || || || || || ||
Db      39583 CTGAGCGTATTACGCAAAACGGCTCACGGGAGAAAAACAGCGCTCGATTATTAGAAAAACTG 39642

QY      490 gacttgcaaaaaaagattctgcaaggctaccttctctctcttctgggataaagagggaatcgc 549
         || || || || || || || || || || || || || || || || || || ||
Db      39643 GACTTTTCATACAAAAGTACAAGAAGTTATCAAGAGCTGATGAAGCGATTTCCTCGGAAAGA 39702

QY      550 attgcaagattgtagctagctctctctctctctctctctggagcaagtgtg 590
         || || || || || || || || || || || || || || || || || || ||
Db      39703 TTTTCAATTCAGTTGATGCGGGCAATCCAAAGACCTCGTCGT 39743

RESULT      9
LOCUS      AP003130
DEFINITION      Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
               genome, section 2/10.
ACCESSION      AP003130 BA000018
VERSION        AP003130.2 GI:14349172
KEYWORDS
SOURCE        Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315,
               strain:N315) DNA.
ORGANISM      Staphylococcus aureus subsp. aureus N315
               Bacteria; Firmicutes; Bacillus/Clostridium group;
               Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE      1 (sites)
AUTHORS        Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
               Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
               Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
               Mizutani-Ui,Y., Kobayashi,N., Sawano,T., Inoue,R., Kaito,C.,
               Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
               Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
               Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
               Whole genome sequencing of methicillin-resistant Staphylococcus
               aureus
TITLE          Lancet 357 (9264), 1225-1240 (2001)
JOURNAL        21311952
MEDLINE        2 (bases 1 to 301050)
AUTHORS        Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K.
               and Kikuchi,H.
DIRECT SUBMISSION
               Submitted (30-JAN-2001) Kenichi Aoki, National Institute of
               Technology and Evaluation, Biotechnology Center; 2Chome 49-10
               Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
               (E-mail:aoki-kenichi2@meti.go.jp, URL:http://www.bio.nite.go.jp/,
               Tel:81-3-3481-1972, Fax:81-3-3481-8424)
COMMENT        On Jun 12, 2001 this sequence version replaced gi:13700170.
FEATURES
               Location/Qualifiers
               source
               1..301050
               /organism="Staphylococcus aureus subsp. aureus N315"
               /strain="N315"
```



```
/transl_table=11
/product="holin-like protein LrgB"
/protein_id="BAB41477.1"
/db_xref="GI:13700179"
/transl_table=11
/transl_table="MINHLALNTPYFGLLSVIPFFLATILPEKTRFFLEFAPLVSM
VGFVAFYLLTGIPYKYGKGGDIYFLEPFAICFAIPLYKKEVLYKWHRIIGIG
IGFVAILLITLFAKLAQFANDVILSMQFQATATLPSAGIGVEKELTSLAVILN
GVIIYALGNKFLKFLRNTPIANGALGTSGHTLGVAPAKELGVESMASIALVLVG
VVVAVVPVFAIFF"
complement(10357..11061)
/gene="SA0254"
CDS
/gene="SA0254"
/notes="ORFID:SA0254
hypothetical protein, similar to transcription regulator
GntR family"
/codon_start=1

Query Match      20.5%; Score 131; DB 1; Length 301050;
Best Local Similarity 55.2%; Pred. No. 7.4e-26;
Matches 323; Conservative 0; Mismatches 250; Indels 12; Gaps 3;

QY 15 tttagtctctttaggaccagaggagcagcgaaccagtgcttttagagcctgct 74
Db 215548 TTTTATACCTTTGAGGCCCGCAGAGCTCTGGAACAACTGTAATTAATGAAGTTA 215607

QY 75 accaatTTtagagaaagagtagaggtgttgacgacctgaacctggcgagctt 134
Db 215608 CCATAGATTA--GTAAAGATTATGATGTCATTATGACATAGACACACGAGTGCTCC 215664

QY 135 gattggggagaagattcgggaagtatttggatcccaagtcatactcagatggatgtaa 194
Db 215665 TACTGGTGAAGAATACTGTAATTTATTAGAGGCAATGAT-----ATGGACATTAG 215718

QY 195 aacagagctacttctctatattgacagtcagacacacatttggtagaaagttcttc 254
Db 215719 AACTGAAGCAATGTTATTTCGCTGCTAGTGAAGAGAACATCTTGTATTAAGGTCATACC 215778

QY 255 agccttgaagctggcgaagttggtcatcattggtctttatcatgattctgttgccta 314
Db 215779 AGCTTTAAAGAGAGTAAGTTGTTGTGTGATCGCTATATCGATAGTTCATTAGTTA 215838

QY 315 tcaggggatttgctggttagatatttgaagccattgactggctcattcagtttgcgac 374
Db 215839 TCAAGGTTATGCTAGAGGGATTGCGTGTGAAGAGTAAGAGCATTAACAGCAATTTGCAAT 215898

QY 375 agatgggctcaaacccgatttgacactctatttgaatcgagtggaagagggctggc 434
Db 215899 AAATGGATTATATCCAGACATGACGATTTATTTAAATTTTAGTCTGAAGTAGTCCGGA 215958

QY 435 tcgtattgctgcaatagtgaccgcgaggttaactctgttgattggaaggggttgactt 494
Db 215959 ACGTATTATT--AAATTCAGAGATCAAAATAGATTAGATCAAGCAATTTAAAGTT 216015

QY 495 gcataaaagtttgcgtaaggtcattctctctcttgcgataagagggaaatcgattgt 554
Db 216016 TCACGAAAAAGTAATTGAAGTTACCAAGAAATCATTCATAATGAATCAACACGCTTCAA 216075

QY 555 caagattgactagtcctccttggagcgaagttgtggaactac 599
Db 216076 AAGCGTTAATGCAGATCAACCTCTTGAATAATGTTGTTGAAGACAC 216120
```

```
RESULT 10
LOCUS AP003359 343590 bp DNA BCT 30-MAY-2001
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 2/9.
ACCESSION AP003359 BA000017
VERSION AP003359.2 GI:14246074
KEYWORDS Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
```

```
ORGANISM Staphylococcus aureus subsp. aureus Mu50
REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group;
AUTHORS Bacillus/Staphylococcus group; Staphylococcus.
1 (sites)
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Ui,Y., Kobayashi,N., Tanaka,T., Sawano,T., Inoue,R.,
Kaito,C., Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S.,
Yabuzaki,J., Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K.,
Yoshino,C., Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and
Hiramatsu,K.
aureus
Whole genome sequencing of methicillin-resistant Staphylococcus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
REFERENCE 2 (bases 1 to 343590)
AUTHORS Ohta,T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@sakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT On May 29, 2001 this sequence version replaced gi:13874621.
FEATURES
Location/Qualifiers
1..343590
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
441..1331
/gene="SAV0307"
441..1331
/gene="SAV0307"
/notes="similar to outer membrane protein precursor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56470.1"
/db_xref="GI:14246075"
/transl_table=11
/translation="MNKISKYIAIASLVAVTSAPOTTNSTAPAKSSAEVQQTQAS
IPASQANLGNQIMAVWYONSAAKALYLOGYNSAKTQLDKREIKNKGHKLIAL
DLDETVDLNSPYQGYASIHNNPPEGHWWQAQAKAPVYGAKEFLYADKGVDIYV
ISDRDEKDLKATQKNLKOQGIQPAKSHILLKGDOKSKESBROMVQKHLMLEFG
DNLLDTDPKEATAESREALIEKHKDDFGKKYIIFPNPMTGSWEATYNNNYKASDKA
KKLRKNAINKQFDPKTGEVK"
1580..2629
/gene="SAV0308"
1580..2629
/gene="SAV0308"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB56470.1"
/db_xref="GI:14246076"
/transl_table=11
/translation="MFLANNEIRNKLKGLIIGVLTMIISVLLFLLSLGLNLMNK
EGIDKQDAIVLNDANQIVQSVFNKKDIEKNYKKAQTLKOTGEIVSNGHOKDNYL
VFGVERSSFLVPSLIEGHRATKDNELATKNGKFGIKDTLSLSQDEKLHIVGT
ESAKYNASPVFTNDATIAKINPLRTGDKINAVVVRDNNKDKLNQDELAIVNDFI
ENLPYGKPNQLNFMISFLVISATVIGFLYVMTLQKTSFLGILKAOQFTNGYLAN
VVISQTLILALFCTAFGLLLTGTGTFDAVPKVPKFDVLLVFAIVLMIVSVLGSFL
SVLTIRKIDPLKAIK"
2642..3319
/gene="SAV0309"
2642..3319
/gene="SAV0309"
/notes="similar to ABC transporter ATP-binding protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56471.1"
/db_xref="GI:14246077"
```

```
/translation="MLKPNVTKSKFGDNRNIEAVKDTNFEINKGDIALLVGPSSGSK
SFTLTAGACTPTSGHILINNODITTMKOKALAKVRMSEIGFILQATNLVPELTVKO
QFTLLKKKNVMNEDYQQLMSQLGTLTLNKLPSSETSGGOKORVAIAKALYTNPSI
ILADEPTAALDTENAIEVIKILRQAKRRKKACIIVTHDERLKAICYDRSYHMKDGVLEN
LENETVE"
3527..4558
/gene="SAV0310"
CDS
3527..4558
/gene="SAV0310"
/note="similar to regulatory protein Pfor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56472.1"
/db_xref="GI:14246078"
/translation="MDLMIGTFLFLVLVITFLTYKAPNMRAMGALANAIATFLV
EAPNKYVGGEVFGKFLIEELGDAAGLGGVAAAGLTALAGSPVYALVIAAAGGMD
LLPFAGYMGYVNMKYTEKYVPDGBLIGSIVILAPLARLIAVLLPVVNSTLRIG
DIQOSTNTNPIIMGIIILGGIITVGTAPUSSMAALTALLGLTGVPMAGAAAFSSAF
MNGTLFHLRLGDRKSTIAVSIIEPLSQADIVSANPIIYITNFFGGAIGLIAMSGL
INDATGTATPIAGFLVMFGFNHPMTIIVYGVMAIVGALAGLYGSIVFKKYPVITKQD
MINRGAVDA"
4901..6019
/gene="SAV0311"
CDS
4901..6019
/gene="SAV0311"
/note="similar to carboxydrate kinase PfkB family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56473.1"
/db_xref="GI:14246079"
/translation="MSDSEKILKRIKDNPPISOELAEALGSRPSVANIISGLIOK
EYVNGKAVLNEDYPIVCIGAANDVRKFYVHKDLVAETSNPTVSTRISGGVARNIAEN
LGRGETVAFLSAGSDSEWENIKRLSTPPNMLDHVOQFENASTGYSYALISKEGDMT
YGLADMEVDFEFLIKRSHLLKAKCIIVDNLKREALNFCAYTHQIKLVIT
TVSSPKMNPDSLSHAIDWIITNKDETEYILNLKIESTDLDLKAARKNNDLGVKNVIV
TNGKELIYRSGEETIKSVMPNSVRKVTGAGDSFCAAVVYSLWNGMSTEDILIAGM
VNAAKTTETKYTVRQNLDOOLYHDMEDYKNGKTKVY"
5994..6917
/gene="SAV0312"
CDS
5994..6917
/gene="SAV0312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB56474.1"
/db_xref="GI:14246080"
/translation="MANLQKYEYSREVOQARENNOPIVALESTIISHGMPYPONVEM
ATTVEQIIRNGATIPATIAIDGKIKIGLESEDLILATSKDVAKVRSDRLAEVAMK
CVGATVTATMICAAMAGIOFFVTGGIGGVHKGAEHTMDISADLEEKSNVTVICAG
AKSLDLPKTMEXILETKGVPIGVYOTNELPAFFTPRESVKLTSSVETPERLADHLTK
QQLNLEGSIYVANIPIPEHALSKAYIEAINEAVVEAENQGIKGDATPFLLGKIVEK
TNGKSLAANKLVENNAALGAKIAVANKLL"
6928..8148
/gene="SAV0313"
CDS
6928..8148
/gene="SAV0313"
/codon_start=1
/transl_table=11
/product="probable pyrimidine nucleoside transport
protein"
/protein_id="BAB56475.1"
/db_xref="GI:14246081"
/translation="MNLIFALTGIATFALFVAFLEFSFRKKNIDFKKLTIMIFIQVILV
FMNMTTGLTILTAFFLEGLINVKAGINVFEGDIQNKNGFTFFLNLLPLVFISV
LIGIFNTIKVLPFIKIVYGAINKIRTWGRLESYFAISTAMFGQPEVYLTIKDIPRL
SRAKLYITAGSGAVSAMGLSYMOMIEPKVYTVAVMLNIFSALITASVINPKSDD
SDVEIDLNTSTETKSVNGTKGKPKVAFQOMIGDSAMDGFKIAVVAVMALLAFISM
EAINMFGSVGLNFKLIGIVFAPIAFLMGPVSEAVPAGSLMATKLTITNEFVAMLOF
KNVLGDSARTQGLIISVLSFANFTVGIIVGIGISDRKQGEKVSFAMRLLGST
complement(8253..9785)
/gene="SAV0314"
complement(8253..9785)
/gene="SAV0314"
/note="similar to sodium-coupled permease"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56476.1"
/db_xref="GI:14246082"
/translation="MKBVGFGLNWAIVITVLLAMLFIGVVFTRKASQSTNSPFTASG
RLPSWVVGFSIYATTLTAITFMSTPEKAFITDWSYIAGNIAIVAIIPLLIVFYVPEFK
KLKTSAYEYLEARPGFSIRVIGSLFLVWHLGRVAIVLYLPTLAITSVSDMNPYIVA
SLVGLLCILYTFLGGFVGWSDSIQGVILGGLAVIILGVNMINKGGFGVVFADAIE
HKLISADNMKLNLTAAAIPIIFLGNFNLLYQYTSQDVVQRYQASDSLKETNKSLLW
TNGLILATSLAPLFYGMGTMLYSFYTHEAVLPKGFNTSSVVPYFILTTPMPFPVAGLLIA
AIPAAASNTSSLSISACISIDIKORFTGKDSERHEVNFARFIIITAGTFFGFMSL
XILPANSNDLMDLFLVFTGLFVPLAGVAVGIFTKRTNTEGVLCGLILGILIFRAVVN
GVGKGNFPFVSTISITVAVFAYILSFIVPSKKHKDITGLTIFEKDKPSTYISKATAT
NK"
complement(9825..10706)
/gene="nanA"
complement(9825..10706)
/gene="nanA"
/note="SAV0315"
/codon_start=1
/transl_table=11
/product="N-acetylneuraminate lyase subunit"
/protein_id="BAB56477.1"
/db_xref="GI:14246083"
/translation="MNKDLKGLYAALLVPDPENGQVNEQGLKQIAQNAITEELDGLY
VNGSSGENFLNTEQKQVFKVAEYGDVKLIAQVSLDNLNIAISIEQFSELPNHEK
ALSAVTPFYPTFEERIDYFDIETATQNNMIIYAIPLDTGVNIAISIEQFSELPNHEK
IVGVKYPAPNFELLRURKAPDPDKLILSGDEMLVQATISGVDGAIGSTYNNVNGRRAR
KIFPLAQGOIQEAYOLQHDSDNIIETVLSMGIYPTLKEILHRHDIDAGLPRKPEKPF
NEARQRLDQLIAKYDL"
10864..11724
/gene="SAV0316"
CDS
10864..11724
/gene="SAV0316"
/note="similar to glucokinase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56478.1"
/db_xref="GI:14246084"
/translation="MYVIAIDIGGTQIKSAVIDQLNMFDYQQIISTPDNKSELITDKV
```

```
Query Match 20.5%; Score 131; DB 1; Length 343590;
Best Local Similarity 55.2%; Pred. No. 7.5e-26;
Matches 323; Conservative 0; Mismatches 250; Indels 12; Gaps 3;

Qy 15 tttagtctcttgaggagcagagggagcagcagaccagtgcttttagaggctcgtc 74
Db 187808 TTTTATAACTTTTGGGGCCAGAGGCTCTGGAACCACTGTAATTAATGAAGTTTA 187867

Qy 75 accaatttagagaaaggagtagaggtgttgacgacccgtgaacctggcgagctct 134
Db 187868 CCATAGATTA---GTAAAGATTATGATGTCATTTAGTACTAGAACCCAGGTGCTCC 187924

Qy 135 gattggggagaaattcgggaagtgtatttggatccaagtcatactcagatggatgctaa 194
Db 187925 TACTGGTGAAGAAATACGTAAATTTGATTAGAGGCAATGAT-----ATGCACATTAG 187978

Qy 195 aacgagagctaactctctctatttggccgacgacagcatttggggaaaaagttctcc 254
Db 187979 AACTGAACCAATGTTATTTGCTCATCTAGAGAGAACAATCTTGTATTAAAGGTCATACC 188038

Qy 255 agccctgaagctggcaagtggcatcatgatgcgtttatcagatgtctctgttcgcta 314
Db 188039 AGCTTTAAAGAGAGTAAGTTGTGTGTGTGATCGCTATATCGATAGTTCAATAGCTTA 188098

Qy 315 tcagggtatttgctggtgcttagatattgaagccaattgactggctcaatcagtttgcgac 374
complement(8253..9785)
```

Db 188099	TCRAGGTTATGCTAGAGGGATTCGGCTTGAGAGAACTGAAGAGCATTAAGAGCATTAACCGAATTTGCAAT	188155	
QY	375	agatgggctcaaacccgatttgacactctattttgacatcgcaggttggaagaagggctgpc	434
Db 188159	AAATGGATTATATCCAGACTTGACGATTATTTAAATGTTAGTCTGGAAGTAGGTGCGCA	188218	
QY	435	tctattgtctgcgaatagtagccgcgcaggtttaatctctgttggatttggaaagggattggaactt	494
Db 188219	ACGTATTATT-- --AAAATTTCAAGAGATCAAAATAGATTAGATCAAGAAGATTTAAAGTT	188275	
QY	495	gcataaaaaagttcgtcaaggtctacacctctctctctgtgataaagaggggaaatcgcatgtg	554
Db 188276	TCACGAAAAAGTAATTGAAGGTTTACCAGAAATCATTTCAATGATCAACACGGTTCAA	188335	
QY	555	caagattgatgctagctctcccttggcagcaagttgtggaactac	599
Db 188336	AAGCGTTAATGCAGATCAACCTCTTGAAATGTTGTTGAAGCAC	188380	
RESULT 11			
LOCUS	MC109	971 bp	DNA
DEFINITION	M.capricoluum DNA for CONTIG MC109.		
ACCESSION	233079		BCT
VERSION	233079.1	GI:541703	
KEYWORDS	atp-binding pyrimidine kinase.		
SOURCE	Mycoplasma capricoluum.		
ORGANISM	Mycoplasma capricoluum		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes; Entomoplasmataceae.		
AUTHORS	1 (bases 1 to 971) Bork,P., Ouzounis,C., Casari,G., Sander,C., Dolan,M. and Gillevet,P.		
TITLE	Exploring the Mycoplasma capricoluum genome: A Parasite reveals it's physiology		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 971)		
AUTHORS	Gillevet,P., Ally,A., Barton,F., Brenner,S.E., Clark-Whitehead,R., Dolan,M., Douglas,N., Hsu,E., Purzycki,M.S., Richer,B., Russo,S., Sartell,J., Smith,S.W., Wang,C., Williams,J. and Gilbert,W.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-SEP-1994) Mycoplasma capricoluum Genome Project, Harvard University, Cambridge, MA 02138 Prokaryotic Genomes Project, Institute for Computational Sciences and Informatics George Mason University, Fairfax, Virginia, USA, 22030-4444 E-mail: gillevet@eranus.ncbgr.nih.gov		
REFERENCE	3 (bases 1 to 971)		
AUTHORS	Bork,P., Ouzounis,C., Casari,G., Schneider,R., Sander,C., Dolan,M., Gilbert,W. and Gillevet,P.M.		
TITLE	Exploring the Mycoplasma capricoluum genome: a minimal cell reveals it's physiology		
JOURNAL	Mol. Microbiol. 16 (5), 955-967 (1995)		

0005304.1

NOTES:

The tga codon is read as Tryptophan in Mycoplasma capricolum Coding sequences below were putatively identified by Peer Bork using the program Genequizz at the EMBL.

EMBL Meyerhofstr 1, 69012 Heidelberg, Germany

E-mail bork@embl-heidelberg.de

More than 870,477 bases have been sequenced by the genomic walking technique and assembled into 372 non-overlapping contigs (Accession numbers Z33005 to Z33376) covering 214,528 base pairs. The length of the contigs varies from 63 base pairs to 2049 base pairs with an average length of 658 base pairs. The data consists of 13,091 bases (68%) from one pass fluorescent sequencing and 201,437 bases (94%) produced by genomic walking. Of the latter, 155,805 bases (77%) have multiple coverage on at least one strand and 112,621 bases (56%) are covered on both strands. There is a total of 968 ambiguous nucleotide assignments in the data set (0.5%). We have compared 8,868 bases of our data with Mycoplasma capricolum sequences already stored in the public databases and note less than 0.7% difference between the two data sets (including ambiguous calls, insertions, deletions and mismatches).

COMMENT

FEATURES
source Location/Qualifiers
l..971
/organism="Mycoplasma capricolum"
/strain="ATCC 27343(Kid)"
/db_xref="taxon:2095"
CDS 84..380
/partial
/note="ORF identified by homology to SwissProt entry gplz19544|ASPEVA240L_1; KTHY_HUMAN; Probability 5.2-05"
/codon_start=1
/transl_table=4
/codon=(seq:"tga",aa:Trp)
/evidence=not_experimental
/product="ATP-bind, pyrimidine kinase"
/protein_id="CAA83746.1"
/db_xref="GI:950071"
/db_xref="SPTREMBL:Q49001"
/translation="-MFITEGMDGSKTALLKVKKEELRLNKVLITREPGEVTAE QIKRIIDLNKKMDAWTALFIARISNOHLQKVIKPALEKNICIFRSLYWFN"
BASE COUNT 429 a 79 c 146 g 315 t 2 others
ORIGIN

Query Match	17.2%;	Score 109.6;	DB 1;	Length 971;
Best Local Similarity	50.3%;	Pred. No. 3.7e-20;		
Matches 294;	Conservative 0;	Mismatches 289;	Indels 1;	Gaps 1;

```

QY      6 aaaaggatttttagctctcttaagggaccagcaggagcgcaagaccagtgttttaga 65
        |||||
Db     77 AATAAATGTTTATTACATTTGAAGGAATGGATGGTAGTGGAANAACACTACTGCTTTATT 136
        |||||

QY     66 ggctctgctaccaattttagagaaaaggaglagagtggttgacgcacctgtgaacctgg 125
        |||||
Db    137 AAAAGTAAAGAACAAGCTAGAAAGCGCTAAATTTATAAAGTTTTTAATTACAGAGAACCTGG 196
        |||||

QY     126 cggagtccttgattggggagaagattcgggaagtgtatttggatcccaagtcatactcagat 185
        |||||
Db    197 TGGAGAAAGTTATTGTCTCAACAATAATTAGGCCAATATTATTTAGATAATAAAAAATAAGATAT 256
        |||||

QY     186 ggatgctaaaaacagagctaactctctatatattgccagtcgcagacagcatitggtggaaaa 245
        |||||
Db    257 GGATGCTTGAACTGAGGCTTTACTTTTTTAGCTTCAGAAATCAACATTTACAAAAGT 316
        |||||

QY     246 agttctccagcccttgaagctggca-agttggtcatactggtcgcttttatcgatagtt 304
        |||||
Db    317 AATAAGCCAGCTTTAGAAAAAATATTATTGTTATTTAGATCGCTTTATTGATTCAA 376
        |||||

QY     305 ctgttgctatcaggggatttggctggtgttagatatattaagccattgactggctcaatc 364
        |||||
Db    377 CTAGTGTCTTATCAAGGAAGTGTAGAAATATTGGTGTGTGATAGTTAGTGAAGTTCAAC 436
        |||||

QY     365 agittgcacagatgggctcaaacccgatttgacactctatttggacatcgaggtggaag 424
        |||||
Db    437 AAATGTGTTTTAAAAAATGTGTTACCCTGATCAACATATTTTTTTGATGTTCTTTTAGTG 496
        |||||

QY     425 aagggtcggctgctattgctgcgaatagtcaccgcgagggttaatcgttggatttgggaag 484
        |||||
Db    497 AAGYTGAANAAGNAATGCAAAATAAGAGAGAGAAAGCTCAAAAACAGACTAGATAAAGAG 556
        |||||

QY     485 ggttggacttgcataaaaaaagtctgcgaaggttacctttctctcttgataaagagggaa 544
        |||||
Db    557 AAAATAACTTTAAACAAAAAGTTTATCAAGGATATTATAGAGCTAGTTAAAAAATAATCCAA 616
        |||||

QY     545 atgcgattgtcaagattgatgcctagctccctcttggagcaagtt 588
        |||||
Db    617 AGCGTATTAAAGTAATTGACGCTCAATAAAGATATTGATCAAGTT 660
        |||||

```

RESULT 12
AX067453 LOCUS AX067453 49617 bp DNA PAT
DEFINITION Sequence 28 from Patent WO0078968.
ACCESSION AX067453


```
NAQAVVLAERYADVHCISGVVHPLDLEPSGAPALDMLLGLLAHPRVVAIGETGLDHYH
EPAAEQAEAFRLHLEAARITAKPVVHTREARADTLALLREAAFLPQAGVLHCFTED
WENARALDGFYISLSGIVTFERNADALREVARQVPADRLLVETDPSYLAIPHRGKP
NIPEYVDVAEFUATLRGVSLETLAEQTTFANFRLFLAR"
complement(2980. .3336)
/gene="pilz"
/note="PA2960"
complement(2980. .3336)
/gene="pilz"
/codon_start=1
/transl_table=11
/product="type 4 fimbrial biogenesis protein pilz"
/protein_id="AAG06348.1"
/db_xref="GI:9949058"
/translacion="MSLPNLGNRNGILSITIKDKSVLYAAVMPFIRNGGLFIPTNKN
YKLGDFVFMLLNMEPEKIPVAGKVWVITPKGAQGNRAAGIGVQVQNDGNTARNKIE
TYLAGALKSDRPTHM"
complement(3370. .4356)
/gene="holB"
/note="PA2961"
complement(3370. .4356)
/gene="holB"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, delta prime subunit"
/protein_id="AAG06349.1"
/db_xref="GI:9949059"
/translacion="MADIYVQOALWSQGGRAQHAHAYLLYGPAGIGKRALAEHWAA
OILCORPAAAGCECKACQLLAAGTHPQVYFVLEPEAEKPIRVQDVRDLVGFVQTA
QUGKRVVLEPAEMVNNANALLKSLEPQSDTVILLISHQPSRLPTIKSRVQOQ
ACPLPQAASLEWALRALDEPAEELLEALLSGGSLPTAQRHLGGVGRVQRAQVVEG
VKRLQQTAAASPLAESWNSVPLPLFDWCFDWTGLIRYQLTHDEGLGLADMRRVI
QYLDKSGQAKVLAMQDMLLOQRQKVLNKANLNLVLEALLVQWASLPQPG"
complement(4349. .4981)
/gene="tnk"
/note="PA2962"
complement(4349. .4981)
/gene="tnk"
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="AAG06350.1"
/db_xref="GI:9949060"
/translacion="MTGLFVTLLEGPEGAGKSTNRDYLAERLRERIEVQLTRPGGTP
LARIPELLASDEPMADTELLMFAARQALAGVIRPALARGAVVLCDRFTDATY
AYOGGGRGLPEARIAALESFVQGLRDLPLVDLPVEIGLAPAAARGRLDRFEQEDR
RFEAVRQTYLQRAQAPERYQVLDAGLPLAEVQAGLDRLLPMLLERLNG"
complement(5010. .6059)
/gene="PA2963"
complement(5010. .6059)
/gene="PA2963"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG06351.1"
/db_xref="GI:9949061"
/translacion="MRKLLVLESGLLFVCLVCGLAAWQOQRALEQPLQTLTEERLLDV
SSGSTPGMLARLEQKVLHGAFWRLRYRPNLPGQALHSGEYRLLPGMKGADLLLELM
REGEVQYSLTVLGESFQVREALARQKLEQTLAGLSDEITMQRLKRPDEVAEGRF
FPDYRTGMRRDIDILRKAYQRMOTILAKEDWGRSODLPYRDAYQALIMASLVERET
GVPEERSQTAGVFRRLQRGMLLQTDPTVIYGMGERYNGKITRADLREPTPYNTYVVP
GMPPTPALAGRAIRAALHPAEGETLYFVARGDGSHTVSSSLDEHNKAVREYQLKRR
SDYRSSPAPITPPQ"
complement(6065. .6880)
/gene="PA2964"
complement(6065. .6880)
/gene="PA2964"
/codon_start=1
/transl_table=11
/product="4-amino-4-deoxychorismate lyase"
/protein_id="AAG06352.1"
/db_xref="GI:9949062"
/translacion="MLDWVDGRPAAELSVDRGLAYGDGLFETLAVRAGTPRLLERHL
```

```
ARLEGCRRLIAIPLDTAALRQELLAFCAALGDGVAKLIVITRGEELRGYAPAPASPRR
ILSGSPRPAYPERHWQGVRLFACTRLAEQPLLAGLKLHNRLEGLARAEWSDAGHA
EGLMLDVHERVQGVSNLLLDVLDGTLVAPDLRRCGVAGVMRAELLERAGGIGVPLAI
RDYSMAELATADEVELCNSQFGIWPVRALDEHWVPVGGELTRKLDQDLRDDLDL"
complement(6880. .8124)
/gene="fabF1"
/note="PA2965"
complement(6880. .8124)
/gene="fabF1"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-acyl carrier protein synthase II"
/protein_id="AAG06353.1"
/db_xref="GI:9949063"
/translacion="MSRRRVITGMGLSPGLDVPSSWEGILAGRSGIAPIEHMDLS
AVSTREGGSKVGFNFVEYLSAKEARLDLPIQXGLAASFOAVRDSGLVENDANRERIG
YSMSGIGGLTIENNCRSLFEQGPRIISPFVPGSLIINNVSGFLSHLGLQPNYAL
TYACTTGTSHIGAARNIATGEADVWAGGSEMAACGLGGGFGGAARLSTRNDEPTR
ASRPDRDRDGFVLSDGSGALVLEELHARAGARIYAEIVGFGMSGDGFHMTAPPED
GAGAACRMKNALRDAGLDPRQVDYINAHGTSPAGDTAEIAAYKSVFGEHAHALSMSS
TKSMTGTHLLCAAGAVEAIFSVLALRDQVAPPTINLDNPDGCDLIDLVAHEAKPRKIDV
ALSNSFGFGTNGTLVFRFAD"
complement(8253. .8489)
/gene="acpP"
/note="PA2966"
complement(8253. .8489)
/gene="acpP"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAG06354.1"
/db_xref="GI:9949064"
/translacion="MSTIERVKKIVAEQGVKEEVNTSASFVEDLGADSLDTVELV
MALEEFETEPDEKAETITVQEAIDYIVAHQO"
complement(8685. .9428)
/gene="fabg"
/note="PA2967"
complement(8685. .9428)
/gene="fabg"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein"
/protein_id="AAG06355.1"
/db_xref="GI:9949065"
/translacion="MSLQGVALTGASRGIGQAIALGLGLGAVVIGTATSSAGAK
IAETLRANGVEGAGLVLDVSSDESVAATLEHQHLOPLIVVNNAGITRONLLVRMK
DDWDFVNVNLSLRSLKXAVRGMTRKARWGRIINIGSVVGANGNAGQNTNAAKAG
LEGFTRALAREVGSRAITVNAVAPGFIDTDMTRLPPEAQREALLQIPLGLRQGAEEI
AKYVGLASDGAAYVTGATVPVNGMYMS"
complement(9451. .10389)
/gene="fabD"
Query Match 16.2%; Score 103.2; DB 1; Length 10471;
Best Local Similarity 51.8%; Pred. No. 3.3e-18;
Matches 234; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
Qy 10 gatttttaactctctttagaggaccagaggagagagagagagagagagagagctgttttagagcgt 69
Db 4975 GGCCTGTCTGTCACCCCTGGAGGCCCGGAGGGCCGCAAGAGCCGACACCAACCGCGATAC 4916
Qy 70 ctgctaccaatttttagaggaagagagagagagagagagagagagagagagagagagagagagag 129
Db 4915 CTGCGCCGAGCCCTGCGCGAGCGCGGCATCGAGGTGCGAGTTCACCCCGAGCCTGGCGGC 4856
Qy 130 gtcttgattggggagagagagagagagagagagagagagagagagagagagagagagagagagat 189
Db 4855 ACGCCTCTCGCCGAGCGCATCCGTGAACCTGCTGCTGCGCGCCAGCAGCAGCGATGCC 4796
Qy 190 gctaaaaacagagctactctctctatattgcccagtcgagagagagagagagagagagagagagag 249
Db 4795 GCCCATACCGAGCTGTTCTGATGTTGCGCGCGCGCCAGCATCTCGCGCGGGGTGCATC 4736
Qy 250 cttccagcccttgaagctggcaagtgtggtcatcatgagatcgtttttatcagatgattctgtt 309
```



```

/product="conserved hypothetical protein"
/protein_id="AAF39304.1"
/db_xref="GI:7190496"
/translatability="MDKDLLENVYRFRFRFFKLSILPALLGLWLFPTDILNVLDT
VILSDRVCGILLILLSALSYPNIIILRLGVFGLWVSFSCSPNFSPLVPAHDSLLGF
ATLAIICLLNRPEDLEVGTIPETSDYNPSSGKGKAVLLFSLWLSQRYLTSIAL
NIADATCCLSFESSILMWIYSMLIVLSLTGGERNHTRPKAVCVATFMLFFALGATL
IAILLQSUTNAGINLIATAPVSFAFFDEIRATWOYLTFQSSDRKKLTKVAFYGS
EYRFSEFWERSVLFPFSKACKQAFGLSPNLVAFACILAIAGFVQINTNLAINPTCR
FVYNNACWFTLVLISFSAKSLRHLRLSLLAAGILLSPVIFHLPLEAKPILLSIVAS
GIAGIGLSIGRL"
complement(1350..1985)
/gene="TC0451"
CDS
complement(1350..1985)
/gene="TC0451"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/translatability="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="AAF39305.1"
/db_xref="GI:7190497"
/translatability="MIPLOKLEQAARSFYDQELLMSRONKLSLQDEIHKHKISLPI
IFPSALMTGALCALCIGTILCFINDLFDYFVFLFPLGILSLAFTALLLIFANKR
QNLVSQLOQVATSCYFSLACSCPEGLSKLVKRLVEFTQDEVLPFGFSKRFFAVL
TLAKPSLLAKESSTKTPPEDEIIEKAFSHIREGLYLSGSKLDHDSOLNQN"
complement(1988..2680)
/gene="TC0452"
CDS
complement(1988..2680)
/gene="TC0452"
/notes="ABC transporter, ATP-binding protein; identified by
match to PFAM protein family HMM PF00005"
/codon_start=1
/translatability="ABC transporter, ATP-binding protein"
/product="ABC transporter, ATP-binding protein"
/protein_id="AAF73555.1"
/db_xref="GI:8163221"
/translatability="MLVVKSLCYAHASHFVEDAAAFAPGOISFILGGSGKTTLF
RIIAGLTSPFISGLVWNGQVQELVAYMOQKEALLPWRVKNIIILLPTELGPBKQKP
LVQDNHFKVIOFQSLSLDREFDELSSGORQVVFAMOCLEPKPILLLDEPFTSLD
SLTKEMLYQDVRLAEEGKAIILASHVDQCLEIGETTFPIRNRKLHPISLQKQGI
AGLIQIKNHLI"
complement(2708..2781)
/gene="trna-ile-1"
/product="trna-ile-1"
complement(2708..2781)
/gene="trna-ile-1"
complement(2786..2858)
/gene="trna-ala-1"
/product="trna-ala-1"
complement(2786..2858)
/gene="trna-ala-1"
complement(3049..3777)
/gene="TC0453"
CDS
complement(3049..3777)
/gene="TC0453"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/translatability="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="AAF39306.1"
/db_xref="GI:7190498"
/translatability="MVFIVIMLSKFYKLSLAVLLINMFSPVVEFAEGTTFIGRMK
SWLSGKTSIPPTEDLQTSATKVSLLSKRYDYTGEGSFVQFPFPPHSGQIVEL
PQSDLAIRDYTYVAETPNDSYTVVSWVEYPEKVIDSKPELNQEGFSGLMILALPESQ
VLYLKATEVOGHKALEFWACEDVDYFGRMLVSVNHTLYQVFMVYKGRSPVELDKENI
FIQSFVKTVKVRNSKMDIRKRVSL"
4045..4872
/gene="TC0454"
CDS
4045..4872
/gene="TC0454"
/notes="similar to PID:557478 GB:AE001273; identified by

```

```

sequence similarity; putative"
/codon_start=1
/translatability="conserved hypothetical protein"
/product="3-deoxy-manno-octulosonate cytidyltransferase"
/protein_id="AAF39307.1"
/db_xref="GI:7190499"
/translatability="MCDYSDLRNFFLGRVYLRGCDVFAFLTSKRKVGILPARGSSRPP
GKPLAKILGTLIORSYENALTSRLDQVVDVATDDQRIQFHVVEFGGLCVMTCECAN
GTRERVAESQHPQAEIVVNIQDEPCLSPSIDGLVEMLEGNPAIDIAITVETVD
PDALLTNHKVKCFDPTGRALYFSSRVPNNKRSPTIYLIHGVAFRRFALSYYVKI
PPSLSLADELEQLRVLESGRSIVVHVQVQATGSPVDYDPEDISKVEYLLCPSKVSF"
4848..6458
/gene="TC0455"
CDS
4848..6458
/gene="TC0455"
/notes="similar to PID:557479 SP:O59321 GB:AE001273;
identified by sequence similarity; putative"
/codon_start=1
/translatability="CTP synthase"
/product="CTP synthase"
/protein_id="AAF39308.1"
/db_xref="GI:7190500"
/translatability="MSFKSIFLTGGVYSSLGKGLTRASLALLERQGLKVMALKLDPY
LVNDPGTMNPEHGEVYVDDGVETDLDLGHYHFFSVQSKHSTATSGGIYTRVITK
ERNGEFLGTVQVIVHTNEIINVIQAEHHRPDVLIIEIGTIGDIESLPLEAVR
QFCEHPQDCLSIHMTVPYVYKSAKEIKTKQHSQNLRSIGISPDVILCRSEVSLT
SEYKRISLFCNVPENAVNAIDLENSIYEMPLLARENIVDFLLSKFSGSPKSLDT
DMONLVTTLCGDNHRLVIGLVCKYLEHKDAYKSVFESLSHASTPANCSEIYVPSPE
SEALTQLOSCDCCLIPGGGTRSGWEGKIAAARYCRDNNTPCFGICLGMALVVEYAR
YALUSPLANSLEMDNPTDPVCMOGODTMIRGTMRMGAYPCQITPKSLAEAYKT
DLVOERHRRHYEVNPAYVLDLQNLGRIVGCPQGDLCETIIEIPNHRMMLGVGFHPEF
LSKLAHPHPLFVKLSAALDYSLEKGRE"
6451..6897
/gene="TC0456"
CDS
6451..6897
/gene="TC0456"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/translatability="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="AAF39309.1"
/db_xref="GI:7190501"
/translatability="MNSMKOEAFVGYGKKRIGLAFASAPILLITLPIGSIINTCSSL
ALTAQALITIKERAVTVVFGNPLPMQKSYASSVQSEIOELAAALVRDMTSLFVILWD
ERLSSAQAEMLKNDGCLSRKQKSSDSLAATLILSSFLDSRKLY"
7013..8536
/gene="TC0457"
CDS
7013..8536
/gene="TC0457"
/notes="glucose-6-phosphate 1-dehydrogenase; identified by
match to PFAM protein family HMM PF00479"
/codon_start=1
/translatability="glucose-6-phosphate 1-dehydrogenase"
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="AAF73556.1"
/db_xref="GI:8163222"
/translatability="MEEIKEMGPTLPACPCVMVITGATGDLTARKLPALYNLTKEG
RLSENVGVGFARPRKSHQEFREMQAIONFSSHEIDIRVWESLEHRIFYQANES
EAGYSLSRFSLESVDQYQTKGNRLFIYLPDPDYFQEIIRNLRHQLFTHEOGAQQP
WSRLIIEKPFQVLDQAEQLQCCIDANINEESYRIDHLYGKRVQNLITIRANTLF
ESCNSQYIDHVIQISYSEISIGISGKNGFFKSGMDVMQHLMLQCLLTMPPESEF
SSAIEKKERIKILKILPIREDYVIRGOYGVQVGVSVSGVEENVDNPSVTVV
AKLFDINPWRKGVFPVLOAKGRLPKRTTDIAVIFKSSYDLFENSENCPLCENDLL
IIRIODEGVALQFNCKVCTNKLVRPVKMDRYSYNTVTPAYPRLLDCITLGR
TLTSTNEEVLASWELSPLEQWSKIRPIPNYIAGSLRQGADELLESDRGRWRSY"
8539..9329
/gene="TC0458"
CDS
8539..9329
/gene="TC0458"
/notes="glucosamine-6-phosphate isomerase, putative;
identified by match to PFAM protein family HMM PF01182"
/codon_start=1

```

```

/transl_table=11
/product="glucosamine-6-phosphate isomerase, putative"
/protein_id="AAE73557.1"
/db_xref="GI:8163223"
/translation="MATLISLNDANRLIAESQEDFLOIACYDWISTANKAIQKRGAF
YVALSGGKTPLOIFQETIVKRAAISDCSKIFVFWGDERASEDEAGSNYLKANDILKW
LRIPDQIQRMDTANPKGDELYENLIREHVPDPIEDMVMLGVGEDGHTLSLFPCTAAL
EEKDRLVVNEVNPQLQTRRMWLTFFPIVRQARHLVAYIQGTAKODLCHKLHLPLGRDTF
PIERVGTPLNPLQWLVSSDCCRAADLADIPAECKLEMF"
gene
complement(9326..10198)
CDS
complement(9326..10198)

```

```

Query Match      13.5%; Score 86.2; DB 1; Length 16207;
Best Local Similarity 55.9%; Pred. No. 2.2e-13;
Matches 185; Conservative 0; Mismatches 143; Indels 3; Gaps 1;

QY 7 aaaggatttttagctctcttgagggaccagagagcagggcagagaccagtggttttagag 66
Db 10831 AATAGAGTGTATTGTAGTAGAGCGGAGAGGTCTGGGAAACCCCAATTTACACAG 10772

QY 67 gctctgtaccatttttagaggaagaggagtagaggtgttgacgacccgtgaacctggc 126
Db 10771 GCACCTTTCTAAACGTTTAAATGGAGAGAGGGAAAGAGGTTGTTTAAACAAGGAGCCAGGA 10712

QY 127 ggaagtcttgattgggggaagattcgcgggaagtgattttgataccaagtctactcagatg 186
Db 10711 GGGAGTGCTCTTGGAGAGCAGCTTCGAGATCTTGTATTGGATGTAACACA---GGAGATA 10655

QY 187 gatgctaaacagagctactctctctatatattgccaagtgcgagacagcattttggtgaaaaa 246
Db 10654 TCTTCCTACGCGGAGTTATTACTGTTCTGCGACGCGCTCAGCACATTCAAGAAAAA 10595

QY 247 gttcttcagcccttgaagctggcaagtgtggtcatcatgatcgtttttatcgatagttct 306
Db 10594 ATATTGCTGCTCTTGAACTCTGGAAAAACAGTTATATGTGACCGTTTTTCATGATCTACT 10535

QY 307 gttgcctatcagggaatttgctcgtggttag 337
Db 10534 ATTGTATACCAAGGAATTGCTGGGGGCTAG 10504

```

Search completed: February 15, 2002, 01:16:16
Job time: 10954 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 00:30:27 ; Search time 41.64 Seconds
(without alignments)
3475.491 Million cell updates/sec

Title: US-09-749-972-1

Perfect score: 639

Sequence: 1 atgtcaaaagatttttagt.....gaatgggttgcccaatga 639

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2.6/prodata/2/ina/5A_COMB.seq.*
- 2: /cgn2.6/prodata/2/ina/5B_COMB.seq.*
- 3: /cgn2.6/prodata/2/ina/6A_COMB.seq.*
- 4: /cgn2.6/prodata/2/ina/6B_COMB.seq.*
- 5: /cgn2.6/prodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2.6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	639	4	US-09-259-109-1
2	68.2	10.7	633	4	US-09-277-565-13
3	36	5.6	7218	1	US-08-232-463-14
4	32.6	5.1	2380	1	US-08-572-951-3
5	31.6	4.9	14602	1	US-08-597-236-1
6	31.6	4.9	14602	1	US-08-746-682A-1
7	30.8	4.8	3025	1	US-08-444-734A-1
8	30.6	4.8	14602	1	US-08-597-236-1
9	30.6	4.8	14602	1	US-08-746-682A-1
10	30.4	4.8	311	1	US-08-594-031-71
11	30	4.7	1282	3	US-08-961-083-171
12	30	4.7	2316	1	US-08-246-403A-7
13	30	4.7	2316	1	US-08-246-403A-10
14	29.6	4.6	2133	4	US-09-187-124-1
15	29.6	4.6	5036	4	US-09-177-349-2
16	29.4	4.6	2517	4	US-08-846-234-4
17	29	4.5	2422	4	US-08-952-365-7
18	29	4.5	3937	3	US-08-586-165-8
19	28.4	4.4	289	4	US-09-007-005-17
20	28.4	4.4	289	4	US-09-244-796-17
21	28.2	4.4	502	4	US-09-085-199B-39
22	28.2	4.4	1419	1	US-07-960-981-4
23	28.2	4.4	1419	5	PCT-US93-09634-4
24	28.2	4.4	8082	1	US-08-306-691B-41
25	28.2	4.4	8082	1	US-08-187-785-1
26	28.2	4.4	8082	5	PCT-US93-06251-28
27	28.2	4.4	4403765	4	US-09-103-840A-2

28	28.2	4.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
29	28	4.4	463	1	US-07-695-564-10	Sequence 10, Appl
30	28	4.4	463	1	US-08-241-387-10	Sequence 10, Appl
31	28	4.4	1524	1	US-07-752-429E-1	Sequence 1, Appli
32	28	4.4	1524	1	US-07-752-428C-1	Sequence 1, Appli
33	28	4.4	1524	1	US-07-752-428C-3	Sequence 3, Appli
34	28	4.4	176373	3	US-09-128-155-17	Sequence 17, Appl
35	27.8	4.4	1010	2	US-08-975-316-57	Sequence 57, Appl
36	27.8	4.4	1915	2	US-08-557-309B-7	Sequence 7, Appli
37	27.8	4.4	1915	3	US-08-834-306-7	Sequence 7, Appli
38	27.8	4.4	1915	4	US-08-953-674A-7	Sequence 7, Appli
39	27.6	4.3	711	4	US-09-004-731-21	Sequence 21, Appl
40	27.6	4.3	711	4	US-08-749-699-21	Sequence 21, Appl
41	27.6	4.3	738	2	US-08-738-462-1	Sequence 1, Appli
42	27.6	4.3	738	5	PCT-US94-07587-1	Sequence 1, Appli
43	27.6	4.3	774	4	US-09-004-731-18	Sequence 18, Appl
44	27.6	4.3	774	4	US-09-004-731-20	Sequence 20, Appl
45	27.6	4.3	774	4	US-08-749-699-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-259-109-1
; Sequence 1, Application US/09259109
; Patent No. 6270762
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Chalker, Alison F.
; APPLICANT: Ingraham, Karen A.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: tdk
; FILE REFERENCE: GM10201
; CURRENT APPLICATION NUMBER: US/09/259,109
; CURRENT FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-259-109-1

Query Match	100.0%;	Score 639;	DB 4;	Length 639;
Best Local Similarity	100.0%;	Pred. No. 4.7e-211;		
Matches 639;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	atgtcaaaagatttttagtctctctgtgaggaccagaggagcaggagcagaccagtggtt	60	
Db	1	atgtcaaaagatttttagtctctctgtgaggaccagaggagcaggagcagaccagtggtt	60	
QY	61	ttagagctctgtaccacatttttagagaaaagagtagaggtgttgacgaccgtgaa	120	
Db	61	ttagagctctgtaccacatttttagagaaaagagtagaggtgttgacgaccgtgaa	120	
QY	121	ccgtgcggagcttctgattgggagaaagatcggaagtgtttggatccaagtcatct	180	
Db	121	ccgtgcggagcttctgattgggagaaagatcggaagtgtttggatccaagtcatct	180	
QY	181	cagatgatgctaaaaacagagctactctcttatattgccagtcgacagacatttggtg	240	
Db	181	cagatgatgctaaaaacagagctactctcttatattgccagtcgacagacatttggtg	240	
QY	241	gaaaaagtcttccagcccttgagctggcaagtgtgtcatcatcgatcgatttatcgat	300	
Db	241	gaaaaagtcttccagcccttgagctggcaagtgtgtcatcatcgatcgatttatcgat	300	
QY	301	agttctgttcctatcagggattgtgtgtggttagattgaagccattgactggtc	360	

	Query Match	5.1%;	Score 32.6;	DB 1;	Length 2380;
	Best Local Similarity	31.7%;	Pred. NO. 0.41;		
	Matches 53;	Conservative 28;	Mismatches 86;	Indels 0;	Gaps
QY	236	tggtagaaaaagtcttcagacccttgaagctggcaagttagctcatcatgatcgctttta	295		
		: : : : : : : : : : : : : : : : : : : :			
Db	1256	TGRTCATATTATGYTCNGNARRTCRAATTNACRAARTCTCTACNAGGCCNCKNGCY	1197		
QY	296	tcatagttctcttgctcatcagggatttgctcgcttgcttagatttgaagccattgact	355		
		: : : : : : : : : : : : : : : : :			
Db	1196	TGRTGNGCDATRTTRTGDDATACNARNACNSNWCNKGCTAYTGCAATNARNCRTTCN	1137		
QY	356	ggctcaatcagtttgccacagatgggctcaaacccgatttgacactc	402		
		: :			
Db	1136	CKRTARTANGCYTNARRTANACNGSNARNARNGCNRTGTGCARTC	1090		

RESULT 5
US-08-597-236-1

```

US-08/597-236-1
/ Sequence 1, Application US/08597236
/ Patent No. 5733765
/ GENERAL INFORMATION:
/ APPLICANT: STINGELE, Franscesca
/ APPLICANT: MOLLET, Beat
/ TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
/ TITLE OF INVENTION: EXOPOLYSACCHARIDES
/ NUMBER OF SEQUENCES: 19
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pennie & Edmonds
/ STREET: 1155 Avenue of the Americans
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/597,236
/ FILING DATE:
/ CLASSIFICATION: 426
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: EP 95201669.9
/ FILING DATE: 20-JUN-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fanucci A., Allan
/ REGISTRATION NUMBER: 30256
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 14602 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 352..1803
/ OTHER INFORMATION: /product= "epsA"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1807..2535
/ OTHER INFORMATION: /product= "epsB"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2547..3239
/ OTHER INFORMATION: /product= "epsC"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 3249..3995
/ OTHER INFORMATION: /product= "epsD"
/ FEATURE:

```

```

NAME/KEY: CDS
LOCATION: 3249..3995
OTHER INFORMATION: /product= "epsd"
FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 4051..4731
; OTHER INFORMATION: /product= "epsE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4898..5854
; OTHER INFORMATION: /product= "epsF"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6425..7540
; OTHER INFORMATION: /product= "epsG"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7736..8212
; OTHER INFORMATION: /product= "epsH"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8221..9192
; OTHER INFORMATION: /product= "epsI"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9285..10364
; OTHER INFORMATION: /product= "epsJ"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10392..11339
; OTHER INFORMATION: /product= "epsK"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11302..12222
; OTHER INFORMATION: /product= "CDS (eps L) covering CDS
; OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 12233..13651
; OTHER INFORMATION: /product= "epsM"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13732..14305
; OTHER INFORMATION: /function= "CDS on the
; OTHER INFORMATION: complementary strand"
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 230..252
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 274..302
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 340..345
;
; US-08-597-236-1

Query Match      4.9%; Score 31.6; DB 1; Length 14602;
Best Local Similarity 52.2%; Pred. No. 2.7;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 497 ataaaaagtcgcaaggctacccttctctcttgataaagggaatcgattgtca 556
      ||||| || | | | | | | | | | | | | | | | | | | | | | |
Db 779 ATAGACACATATCGAGATCTTGATGTCAGCTCTCAAAAAGATAAAAAGTTGATGTTA 838

QY 557 agattgatgctagctccctttggagcaagttgtggaaactaccaaggtctgtgttg 616
      | ||||| | | | | | | | | | | | | | | | | | | | | |
Db 839 AAGTTGATGATGTTGCTCATATCAAGAAGCTTATGATATCTCAAGCTCTGGCAATCTA 898

QY 617 acggaatgggcttg 630
      | | |||| | || |
Db 899 AAGCTATGGTCTTG 912

RESULT 6
US-08-746-682A-1

; Sequence 1, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746.682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 352..1803
; OTHER INFORMATION: /product= "epsA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..2535
; OTHER INFORMATION: /product= "epsB"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2547..3239
; OTHER INFORMATION: /product= "epsC"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3249..3995
; OTHER INFORMATION: /product= "epsD"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4051..4731
; OTHER INFORMATION: /product= "epsE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4898..5854
; OTHER INFORMATION: /product= "epsF"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6425..7540
; OTHER INFORMATION: /product= "epsG"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7736..8212
; OTHER INFORMATION: /product= "epsH"
; FEATURE:
; NAME/KEY: CDS

```

LOCATION: 8221..9192
OTHER INFORMATION: /product= "epsi"
FEATURE:
NAME/KEY: CDS
LOCATION: 9285..10364
OTHER INFORMATION: /product= "epsj"
FEATURE:
NAME/KEY: CDS
LOCATION: 10392..11339
OTHER INFORMATION: /product= "epsk"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 11302..12222
OTHER INFORMATION: /product= "CDS (eps L) covering CDS
OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
FEATURE:
NAME/KEY: CDS
LOCATION: 12233..13651
OTHER INFORMATION: /product= "epsm"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 13732..14305
OTHER INFORMATION: /function= "CDS on the
OTHER INFORMATION: complementary strand"
OTHER INFORMATION: product= "orfz"
FEATURE:
NAME/KEY: terminator
LOCATION: 230..252
FEATURE:
NAME/KEY: promoter
LOCATION: 274..302
FEATURE:
NAME/KEY: RBS
LOCATION: 340..345
US-08-746-682A-1

Query Match 4.9%; Score 31.6; DB 1; Length 14602;
Best Local Similarity 52.2%; Pred. No. 2.7;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 497 ataaaaagtgctcaagctacccttctctctgataaaagggagaaatcgattgtca 556
|||||
Db 779 ATAAAGCAATATCGAGATCTGTGATCGACTCTCAAAAAGATAAAAAAGTTGATGTTA 838
|||||
QY 557 agattgatctagctccctttggagcaagttgtgaaactaccagcgtctcttctttg 616
|||||
Db 839 AAGTTCATGATGTGCTTCATATCAAGAGCTTATGATATCTCAAGTCTGGCAAAATCTA 898
|||||
QY 617 acggaatgggcttg 630
|||
Db 899 AAGCTATGGTCTTG 912
|||

RESULT 7
US-08-444-734A-1
; Sequence 1, Application US/08444734A
; Patent No. 5610282
; GENERAL INFORMATION:
; APPLICANT: Sibley, David R.
; APPLICANT: Monsma, Frederick J.
; APPLICANT: Mahan, Lawrence C.
; APPLICANT: McVittie, Loris D.
; TITLE OF INVENTION: cDNA encoding the rat D1 dopamine
; TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and
; TITLE OF INVENTION: expression of the receptor protein in plasmid-transfected
; NUMBER OF INVENTION: cell lines
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA

COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,734A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/029,917
FILING DATE: 03-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/548,714
FILING DATE: 06-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH065.001FW1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3025 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: PB73D1
FEATURE:
NAME/KEY: CDS
LOCATION: 293..1756
US-08-444-734A-1

Query Match 4.8%; Score 30.8; DB 1; Length 3025;
Best Local Similarity 53.3%; Pred. No. 2;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 442 gctgcgaatagtcacgcaggttaatcgcttgattgaaagggttgacttgcataaa 501
|||||
Db 2069 GGTGCTAAACTGCTCCGAGCAGTTTTCAGATTGGGAAGGTAGGTGATGCTTTGTTAAT 2128
|||||
QY 502 aaagttcgtcaaggctacaccttctctctgataaaagagggaatcgcttcaagatt 561
|||||
Db 2129 TAACCTTCCTCAATAATAATTGAGCCTTACAGCAGGAGTGGGATTCCTTTTCTCAGAAAT 2188
|||||
QY 562 ga 563
|||
Db 2189 GA 2190
|||

RESULT 8
US-08-597-236-1/c
; Sequence 1, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Franscesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York


```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 352..1803
;   OTHER INFORMATION: /product= "epsA"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1807..2535
;   OTHER INFORMATION: /product= "epsB"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2547..3239
;   OTHER INFORMATION: /product= "epsC"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 3249..3995
;   OTHER INFORMATION: /product= "epsD"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 4051..4731
;   OTHER INFORMATION: /product= "epsE"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 4898..5854
;   OTHER INFORMATION: /product= "epsF"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 6425..7540
;   OTHER INFORMATION: /product= "epsG"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 7736..8212
;   OTHER INFORMATION: /product= "epsH"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 8221..9192
;   OTHER INFORMATION: /product= "epsI"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 9285..10364
;   OTHER INFORMATION: /product= "epsJ"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 10392..11339
;   OTHER INFORMATION: /product= "epsK"
;   FEATURE:
;   NAME/KEY: misc.feature
;   LOCATION: 11302..12222
;   OTHER INFORMATION: /product= "CDS (eps L) covering CDS
;   OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 12233..13651
;   OTHER INFORMATION: /product= "epsM"
;   FEATURE:
;   NAME/KEY: misc.feature
;   LOCATION: 13732..14305
;   OTHER INFORMATION: /function= "CDS on the
;   OTHER INFORMATION: complementary strand"
;   OTHER INFORMATION: /product= "orfz"
;   FEATURE:
;   NAME/KEY: terminator
;   LOCATION: 230..252
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: 274..302
;   FEATURE:
;   NAME/KEY: RBS
;   LOCATION: 340..345
;   US-08-746-682A-1

```

```

Query Match      4.8%; Score 30.6; DB 1; Length 14602;
Best Local Similarity 65.2%; Pred. No. 6;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 493 ttgcataaaaaagtgctcaaggctaccctttctctcttcggtataaaggaggaatcgcat 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10542 TTAATAAAAAAATCTTGAATCTACTTTTTTATCTATATGAAGAAAGAAATCAAAGT 10483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 553 gtcaagatt 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10482 ATTGAGAAAT 10474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-594-031-71/c
; Sequence 71, Application US/08594031
; Patent No. 5783182
; GENERAL INFORMATION:
;   APPLICANT: THOMPSON, Timothy C.
;   TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES
;   NUMBER OF SEQUENCES: 175
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BAKER & BOTTS, L.L.P.
;   STREET: 1299 Pennsylvania Avenue, N.W.
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20004-2400
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/594,031
;   FILING DATE: 30-JAN-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/006,838
;   FILING DATE: 16-NOV-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Remenick, James
;   REGISTRATION NUMBER: 36,902
;   REFERENCE/DOCKET NUMBER: 0A146-0110
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-639-7700
;   TELEFAX: 202-639-7890
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 71:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 311 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: CDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
;   US-08-594-031-71

Query Match      4.8%; Score 30.4; DB 1; Length 311;
Best Local Similarity 51.5%; Pred. No. 0.69;
Matches 70; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 294 tatcgatagttctgttcctcatcaggatttgctcggtcttagattgaagcattga 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 TATCGATGTTGGGTTCTCCTAGCAGGCTCTGGATAGACGACAGACTACTGACATCAT 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 354 ctggctcaatcagtttgcgcagatgggtcacaacccgatttgacactattttgacat 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Search completed: February 15, 2002, 01:49:59
Job time: 4772 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 22:03:21 ; Search time 1181.61 Seconds
(without alignments)
5811.183 Million cell updates/sec

Title: US-09-749-972-1

Perfect score: 639

Sequence: 1 atgtcaaaagatttttagt.....gaatggcgttgccaaatga 639

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:**
2: em_esthum:**
3: em_estin:**
4: em_estom:**
5: em_estpl:**
6: em_estba:**
7: em_estro:**
8: em_estov:**
9: em_htc:**
10: gb_est1:**
11: gb_est2:**
12: gb_htc:**
13: gb_gss:**
14: em_gss_fun:**
15: em_gss_hum:**
16: em_gss_inv:**
17: em_gss_pln:**
18: em_gss_pro:**
19: em_gss_rod:**
20: em_gss_vrt:**
21: em_gss_other:**
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51.4	8.0	251	10	AU063540
C 2	42.6	6.7	1050	13	CNS012BP
C 3	38.6	6.0	682	10	AL509002
C 4	38.6	6.0	1092	13	CNS04YG5
C 5	38.4	6.0	1101	13	CNS00LT2
C 6	36.8	5.8	646	13	A2814364
C 7	36.2	5.7	707	11	BF158171
C 8	36	5.6	535	11	BG691744
C 9	35.6	5.6	562	10	AU154589
C 10	35.6	5.6	758	13	AQ858832
C 11	35.4	5.5	299	13	AQ882061
C 12	35.4	5.5	464	11	BG566008

13	35.4	5.5	480	10	AI021635
14	35	5.5	1101	13	CNS0106X
C 15	34.8	5.4	516	10	AW306038
C 16	34.8	5.4	521	10	BE201309
C 17	34.8	5.4	839	13	CNS03L4D
C 18	34.6	5.4	608	10	AU039978
C 19	34.6	5.4	682	10	AU039910
C 20	34.6	5.4	689	11	C90165
C 21	34.4	5.4	364	11	BG979728
C 22	34.4	5.4	401	10	AA549022
C 23	34.4	5.4	613	11	BF231990
C 24	34.4	5.4	619	10	AA546725
C 25	34.2	5.4	878	13	CNS05EJI
C 26	34.2	5.4	1008	11	BF259733
C 27	34.2	5.4	1201	13	CNS015W6
C 28	34	5.3	562	11	BI425539
C 29	33.8	5.3	474	11	BG578759
C 30	33.8	5.3	515	10	AI727581
C 31	33.8	5.3	595	10	AW640503
C 32	33.8	5.3	599	11	BG162656
C 33	33.8	5.3	618	13	A2874122
C 34	33.8	5.3	646	10	AI727515
C 35	33.8	5.3	1101	13	CNS000D1
C 36	33.6	5.3	470	10	AA050890
C 37	33.6	5.3	775	13	A2127961
C 38	33.4	5.2	580	10	AI881565
C 39	33.4	5.2	859	13	CNS004YY
C 40	33.2	5.2	454	13	A2049503
C 41	33.2	5.2	632	10	AU221260
C 42	33.2	5.2	727	11	C89525
C 43	33.2	5.2	924	11	BF535569
C 44	33.2	5.2	1072	13	CNS0390U
C 45	33	5.2	357	10	AW687837

ALIGNMENTS

```
RESULT 1
LOCUS AU063540 251 bp mRNA EST 30-APR-2001
DEFINITION AU063540 Rice callus Oryza sativa cDNA clone C62637, mRNA sequence.
ACCESSION AU063540
VERSION AU063540.2 GI:13895842
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 251)
Sasaki,T. and Minobe,Y.
Rice cDNA from callus
Unpublished (1994)
On May 20, 1999 this sequence version replaced gi:4879241.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.
C62637_12Z.
Location/Qualifiers
1..251
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C62637"
/clone_lib="Rice callus"
Note=Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
```

oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid.

```

BASE COUNT      36 a      79 c      82 g      54 t
ORIGIN

Query Match
Best Local Similarity 8.0%; Score 51.4; DB 10; Length 251;
Matches 121; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 107 tgacgacccgtgaacctgacgagcttgattgggagaaattcgggaagtgtattgg 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8 TCATGACTCGCGAGCCGCGGATTCCTGCTGGTGGCAGCGATCCGCGAGTGGTCTCT 67

Qy 167 atcaagtcatactcagatgagtgataaaacagagctactctctatatattccagtcga 226
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 CTCCTGGCTCGCGCGGATCTCCTCGCGCTGAGGCTTTCGTGTACAAACCCGACAAGG 127

Qy 227 gacagcatttggtgaaaagtcttccagcccttgaagctggaagtgtgtcatctg 286
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 128 CTCACACGTCGACGAGGTTGTCATTCGGCTTTCGAGAGGCGCAAGCTCGCTTTCG 187

Qy 287 atcgctttatcgatgattctgtgctatcagaggttgctgctggcttaatatattg 343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 ACCGTTACTGACTCGACTATCGCTATCAGGGGCTGGCAGGCCCTCGACCCCTG 244

RESULT 2
CNS012BP/c
LOCUS
DEFINITION CNS012BP 1050 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN07P18 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL101359
VERSION AL101359.1 GI:5612970
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1050)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
Location/Qualifiers
1..1050
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN07P18"
/note="end : SP6"
BASE COUNT 231 a 281 c 64 g 199 t 275 others
ORIGIN

Query Match
Best Local Similarity 6.7%; Score 42.6; DB 13; Length 1050;
Matches 80; Conservative 56; Mismatches 113; Indels 0; Gaps 0;

Qy 295 atcgatgattctgtgctatcagggattggtgtgtgttagatattgaagccattgac 354
      |:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db 919 AKATKWTCTKDGDRKATWAGRGTTTRAWAKKARGDTTTARAARAGRGRTWATRKAR 860
Qy 355 tggctcaatcagatttgcacagatgggtctcaaacccgatttgacactctatttgacatc 414
      :|:|:|||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 859 RRGRTTDAMARGTTTKTRRAGDWTTKGAAAAAATATTTTAAAGAGAAATARAKARATW 800
Qy 415 gaggtgggaagagggtgctgctgatttctgctggaatagtcacccgaggttaattgttg 474
      ||:|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 799 TAGRGCGWRGRRGATGARAGTGTGCTGGGRRGGGRRGGGGRADKGGGKGRG 740
Qy 475 gatttgaaggttggacttgcataaaaaaagtctgaaggtcactctctctctcttgat 534
      ||:|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 739 AGDKRRKKRGRKRRATTTGTTGGRWAAGGTGDTTTTGTAAAGTTRRKGVGAGAA 680
Qy 535 aaagaggga 543
      :|:|:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 679 RTTRMGGW 671

RESULT 3
AL509002/c
LOCUS
DEFINITION AL509002 682 bp mRNA EST 04-JAN-2001
Hordeum vulgare cDNA clone HY10G22V 5', mRNA sequence.
ACCESSION AL509002
VERSION AL509002.1 GI:12035505
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 682)
Michalek W., Weschke W., Pleissner K.-P. and Graner A.
EST sequencing and analysis in barley
Contact: Michalek W
Unpublished (2000)
Institute for Plant Genetics and Crop Plant Research
Corrensstr.3, D-06466 Gatersleben, Germany
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
Seq primer: T3 primer for 5' end.
FEATURES
source
Location/Qualifiers
1..682
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY10G22V"
/clone_lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="XLOLR"
/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
of spring barley variety 'Barke', a high quality malting
variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
(3'-end of cDNA). NOTE: Due to a cloning artefact caused
by the kit, in most cases the EcoRI site is NOT present,
as well as the EcoRI adapter. Average insert size is 1 kb
Sequence trimming: Vector sequences and sequence ends were
trimmed from the 5'- and 3'-end until a 50 bp window
contains less than two ambiguities. The maximum length was
set to 700 bp"
BASE COUNT 160 a 186 c 175 g 159 t 2 others
ORIGIN

Query Match
Best Local Similarity 6.0%; Score 38.6; DB 10; Length 682;
Matches 202; Conservative 0; Mismatches 224; Indels 9; Gaps 2;

Qy 190 gctaaacagagctactctctctattgcccagtcgacagcattgttggtggaagaagt 249
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 587 GATAAAGCCGAAGTTATGTTTATGCGCGCGCTTCAACTGGTGAAGAACGGTCATC 528

```


[illegible]

RESULT	6				
AZ814364/c					
LOCUS		AZ814364	646 bp	DNA	GSS
DEFINITION		2M0081J23R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUGC2M0081J23 R, DNA sequence.			
					20-FEB-2001

VERSION	AZ814364.1	GI:12984188
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 646) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., M., Rose,M., Rose,R., Mahmoud,M., Meenen,E., Pedersen,T., Reilly Islam,H., Longacre,S., Wright,D., Weiss,R., Tingey,A., von Niederhausern,A.	
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb Plasmid inserts	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Robert B. Weiss University of Utah University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E.,	

```

FEATURES
    source
        high quality sequence stop: 040.
        Location/Qualifiers
            1. .646
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC2M0081J23"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a

```

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	170 a	161 c	132 g	183 t
ORIGIN				
Query Match		5.8%;	Score 36.8;	DB 13; Length 646;
Best Local Similarity		44.5%;	Pred. NO. 5.8;	

QY	1	atdcaaaaggatttttagtctctcttgaggaccagagggacgagcaagaccagtggt	60
Db	591	AAGTCAGTATTGTGTGCAGCAAGATGGAGGGAAATTCACAGCAGTGGATTTCTCTGGT	532
QY	61	ttagaaggctctgtaccaaatttttagagggaaaaaggagtagagtggttgacgacccctgaa	120
Db	531	CCTGAGCCTCTGGTGGCAGACGCTGTTGAAGTGGGCAAGAGCAGCGCATGATTCGTTAA	472
QY	121	ctctggcgaggtcttgattggggagagaagattcgggaaagtgaatttggatccaagtcatact	180
Db	471	ACTTTAATGTGATTACATGCGCTCCAATGCTCTCTAGAGGCTTGTCTGATPACAGCTTCGT	412
QY	181	cagatggatgtctaaacagagctactctctatatatgccagtcgcagacagcagcttgggtg	240
Db	411	CTTAAGGTAGGGCTGCAACATTAACCTCGAATTTTGGATACCAGGACACAGCTTGTACTTT	352
QY	241	gaaaaagtctcttcagcccttgagcttgcgaagtggctcatcgatcgattttatcgcgt	300
Db	351	TAAAAAGAAATCGCGGCTTGGAGCTGGGGATATTTTGTAGTTGGTAGAATCGTCCCT	292
QY	301	agttctgttgccctatcagggatttggtc	328
Db	291	AGTACCCTTTGAAGCCTCTGGGATTGATC	264

RESULT	7			
		BF158171/c		
LOCUS				
		BF158171	707 bp	mRNA EST 23-FEB-2001
DEFINITION		fl29e03.y1 Sugano Kawakami zebrafish DRB Danio rerio cDNA clone 3815141 5' similar to TR:Q9Y3W8 Q9Y3W8 HYPOTHETICAL 71.4 KD PROTEIN ; mRNA sequence.		

ACCESSION NUMBER	VERSION	KEYWORDS	SOURCE	ORGANISM
BF158171				
BF158171.1		GI:11053375		
EST.				
zebrafish.				
Danio rerio				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;				
Cypriniformes; Cyprinidae; Rasbora; Rasbora;				

REFERENCE
AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Willson, R.

1/F (bases 1 to 707)

1/F (bases 1 to 707)

TITLE	WashU Zebrafish EST Project 1998
JOURNAL	Unpublished (1998)
COMMENT	Contact: Stephen L. Johnson Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA


```

FEATURES
source
1. 562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4001564"
/clone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
167 a 135 c 88 g 166 t 6 others
BASE COUNT
ORIGIN

```

	Query Match	5.6%	Score 35.6;	DB 10;	Length 562;
	Best Local Similarity	47.1%;	Pred. No. 12;		
	Matches 104;	Conservative 0;	Mismatches 117;	Indels 0;	Gaps 0;
QY	404	attttgacatcgagggtggaagaagggctgctcgtattgctgcgaatatgtagccgcgagg	463		
Db	237	ATATTGATGAAGATGGCGACAGATATTTGGAGATTCGGCTTAAGCAGCGCTGGTGAAGATG	178		
QY	464	ttaatcgtttggtatttggaagggttgagacttgcataaaaaagtttcgtaagggtcaccttt	523		
Db	177	GAGATGATGAAGATTTGGGAAGAAGATGATGCTGAANAAGACTGCTCTGGAAGGCTATTTCN	118		
QY	524	ctctctcgtgataagagggaatacgcatttctcaagattgtagtctgctcccccttggagc	583		
Db	117	CAATCATTGATGATGAAGATAACCCCTGTCATGAGTATACAGATATTTAAAGCATCTCTTTC	58		
QY	584	aagtttggaactaccgaaggtcgtcttgtttgacggaatg	624		
Db	57	NAACTATTCAAAATCGTAATCTCTGTGTGGTATCAGGCACATG	17		

```

RESULT 10
AQS58832/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 758)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence start: 31
High quality sequence stop: 503.
Location/Qualifiers
1..758
/oranism="Oryza sativa"
FEATURES
source

```

```

/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
./clone="nbcb0003B09r"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocytledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

	Query Match	5.6%;	Score 35.6;	DB 13;	Length 758;
	Best Local Similarity	52.7%;	Pred. No. 13;		
	Matches 77;	Conservative 0;	Mismatches 69;	Indels 0;	Gaps 0;
QY	145	agatttcgggaagtgtattttggatccaagtctacatcagatgagtgctaaaaacagagcta	204		
Db	667	AAATTTATGGAAGGATTTGGGACCATTGTCTTGATAAAGTGGGGATTCCAAACAAA	608		
QY	205	cttctctatatttcgactgacagacagcattttggtggaaaaagtcttccagcccttgaa	264		
Db	607	GTTAGCTCAATTGGCAATTACAACAAAAAATTTCTGCTACACACAGCTCTGAAGTGTT	548		
QY	265	gctggcaagtgtggtcatcatcattggatcg	290		
Db	547	TGTGGCCAAATGGGTGTCAAGAATGG	522		

RESULT	11
AQ882061/c	
LOCUS	299 bp DNA GSS 09-NOV-1999
DEFINITION	HS_5337 Al_E11 SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=9105 Col=21 Row=I, DNA sequence.
ACCESSION	AQ882061
VERSION	AQ882061.1 GI:6313528
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 299)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA

ility sequence stop: 173.

```

/clone_lib="Soares_mammary_gland_NbMWG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAATGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
111 a 130 g 98 t
BASE COUNT

```

[illegible]

RESULT	14	
CNS0106X		
LOCUS		
DEFINITION	CNS0106X 1101 bp DNA 26-JUL-1999	
	Drosophila melanogaster genome survey sequence T7 end of BAC	
	BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit	
	fly), genomic survey sequence.	
ACCESSION	AL098595	
VERSION	AL098595.1	
KEYWORDS	GI:5610206	
SOURCE	GSS.	
	fruit fly.	

ORGANISM	Plasmid <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE AUTHORS	1 (bases 1 to 1101) Genoscope.

TITLE	COMMENT
<p>Direct Submission</p> <p>Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191006 EVRY cedex - FRANCE (E-mail : seqrf@genoscope.cns.fr</p> <p>- Web : www.genoscope.cns.fr)</p>	<p>Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (DrosBAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBel0BAC11.</p>

```

FEATURES
source
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"

```

```

/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03K20"
/note="end : T7"

```

BASE COUNT	258 a	107 c	60 g	175 t	501 others
ORIGIN					
Query Match		5.5%	Score 35;	DB 13;	Length 1101;
Best Local Similarity	15.1%		Pred. NO. 22;		
Matches 52:	Conservative 150;	Mismatches 143;	Indels 0;	Gaps 0;	

[illegible]

RESULT	15
AW306038/c	
LOCUS	AW306038 516 bp mRNA
DEFINITION	fj79b05.yl zebrafish gridded kidney Danio rerio cdna 5', mRNA sequence.
	EST 20-JAN-2000

ACCESSION AW306038
VERSION AW306038.1
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 516)

AUTHORS

Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Washu Zebrafish EST Project 1998

**JOURNAL
COMMENT**

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) RessourcenzentrumPrimärDatenbank, Berlin, Germany (web address: www.rzpd.de)


```

Seq primer: T3 ET from Amersham
High quality sequence stop: 433.
Location/Qualifiers
1. .516
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish gridded kidney"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOLR"
/notes="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
Site_2: XhoI; Oligo dt cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."
BASE COUNT      138 a      124 c      126 g      127 t      1 others
ORIGIN

```

```

Query Match      5.4%; Score 34.8; DB 10; Length 516;
Best Local Similarity 50.6%; Pred. No. 20;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 27  tgaggaccagagagcaggcaagaccagtgttttagaggctctgtaccatttaga 86
    || || || || || || || || || || || || || || || || || || ||
Db 203  TGCAGGCCACAGAGCTCCAGAGCTGAGGTCGGGTCGAGTCTCTCTCTGTGTGT 144

QY 87  ggaagaggagtagaggtgttgacgacccgtgaacctggcggagctcttgattggggagaa 146
    || || || || || || || || || || || || || || || || || || ||
Db 143  TGAGCAAGAGAGCTGATGCTGATGTAACTGCTCCAGTTGGACTGATGCTGAGGAGGG 84

QY 147  gattcgggaagtattttgatccaaagtcaactatcagatggatgct 192
    || || || || || || || || || || || || || || || || || || ||
Db 83  CCAGCAGGTATGGCCTGTATGAGGAGGCCACACTCACACCAGTCT 38

```

Search completed: February 15, 2002, 00:50:14
Job time: 10013 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 23:40:32 ; Search time 102.36 Seconds
(without alignments)
5352.001 Million cell updates/sec

Title: US-09-749-972-1
Perfect score: 639
Sequence: 1 atgtcaaaaggatttttagt.....gaatgggttgcccaatga 639

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_1101:*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	639	100.0	639	21	AAA64398	DNA encoding a thy
2	635.8	99.5	639	21	AA291859	Streptococcus pneu
3	635.8	99.5	28171	19	AAV52155	Streptococcus pneu
4	634.2	99.2	1632	19	AAZ96394	S. pneumoniae derl
5	131	20.5	840	21	AAZ51229	Staphylococcus aur
6	125.8	19.7	5855	20	AAZ13042	Enterococcus faeca
7	119.4	18.7	4029	18	AAV74612	Staphylococcus aur
8	107.2	16.8	49617	22	AAF28541	Genomic fragment #
9	89.6	14.0	1038602	20	AAZ01425	Complete genome se
10	86.6	13.6	8102	20	AAZ13181	Enterococcus faeca
11	81.6	12.8	349980	21	AAF21608	Neisseria meningit

12	68.2	10.7	633	20	AAZ37114	Nucleotide sequenc
13	56.6	8.9	58073	18	AAZ58840	Mycoplasma genital
14	42.4	6.6	3332	22	AAH54919	S. epidermidis gen
15	39.4	6.2	936	22	AAF58252	Oligonucleotide D1
16	39.4	6.2	936	22	AAF58254	Oligonucleotide D1
17	39.4	6.2	936	22	AAF58257	Oligonucleotide D1
18	39.4	6.2	936	22	AAF58259	Oligonucleotide D2
19	39.4	6.2	936	22	AAF58262	Oligonucleotide D2
20	39.4	6.2	936	22	AAF58255	Oligonucleotide D1
21	39	6.1	936	22	AAF58252	Oligonucleotide D1
22	39	6.1	936	22	AAF58254	Oligonucleotide D1
23	39	6.1	936	22	AAF58257	Oligonucleotide D1
24	39	6.1	936	22	AAF58259	Oligonucleotide D2
25	39	6.1	936	22	AAF58262	Oligonucleotide D2
26	39	6.1	938	22	AAF58255	Oligonucleotide D1
27	35.4	5.5	805	22	AAH05496	Human cDNA clone (
28	35.4	5.5	1793	22	AAH15903	Human cDNA sequenc
29	33.8	5.3	194	22	AAI21374	Probe #11307 for g
30	33.8	5.3	194	22	AAI46655	Probe #15341 used
31	33.8	5.3	194	22	AAI07069	Probe #7060 used t
32	33.8	5.3	847	22	AAH08614	Human cDNA clone (
33	33.8	5.3	2291	22	AAH16769	Human cDNA sequenc
34	32.8	5.1	32768	20	AAZ12954	Enterococcus faeca
35	32.8	5.1	62909	22	AAF28545	Genomic fragment #
36	32.6	5.1	465	22	AAF94354	Fusarium venenatum
37	32.6	5.1	583	21	AAF09313	Haemophilus influe
38	32.6	5.1	759	20	AAZ16593	Human gene express
39	32.6	5.1	2380	19	AAV66834	Zea mays soluble s
40	32.6	5.1	2478	18	AAV67287	Soluble starch syn
41	32.2	5.0	1605	21	AAZ42127	Arabidopsis thalia
42	32.2	5.0	3263	22	AAH16211	Human cDNA sequenc
43	32	5.0	3333	21	AAZ59415	Human secreted pro
44	31.6	4.9	14602	18	AAZ73236	Streptococcus ther
45	31.6	4.9	15449	21	AAZ81526	N. meningitidis pa

ALIGNMENTS

RESULT 1
AAA64398
ID AAA64398 standard; DNA; 639 BP.
XX
AC AAA64398;
DT 20-DEC-2000 (first entry)
XX
DE DNA encoding a thymidylate kinase polypeptide.
XX
KW Thymidylate kinase; tdk polypeptide; bacterial infection; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
CDS FT 1..639
FT /*tag= a
FT /product= "thymidylate kinase"
XX
FN WO2000050602-A1.
XX
PD 31-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04238.
XX
PR 26-FEB-1999; 99US-0259109.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Burnham MKR, Zalacain M, Biswas S, Chalker AF, Ingraham KA;
PI Trainl CM, Warren PV;
XX
DR WPI; 2000-533181/48.
DR P-PSDB; AAB08637.

QY 241 gaaaaagtcttccagcccttgaaagctggcaagttggtgcatcgtggtatggtttatcgat 300
Db 241 gaaaaagtcttccagcccttgaaagctggcaagttggtgcatcgtggtatggtttatcgat 300
QY 301 agttctgtccctacaggaattggtcgtggttaagatatgaagccattgactggtc 360
Db 301 agttctgtccctacaggaattggtcgtggttaagatatgaagccattgactggtc 360
QY 361 aatcagtttgcagagatgggtcctcaaacccgatttgacactctatttgacatcgaggtg 420
Db 361 aatcagtttgcagagatgggtcctcaaacccgatttgacactctatttgacatcgaggtg 420
QY 421 gaagaagggtggtcgtatgtgtgcaatagtagcgcgaggttaatcgtttgattg 480
Db 421 gaagaagggtggtcgtatgtgtgcaatagtagcgcgaggttaatcgtttgattg 480
QY 481 gaaggttgacitgcataaaaaagttcgtcaaggctaaccttctctctctggataaagag 540
Db 481 gaaggttgacitgcataaaaaagttcgtcaaggctaaccttctctctctggataaagag 540
QY 541 ggaatcgcatgttcaagattgtagtctcctcttgagcaagttgtggaactacc 600
Db 541 ggaatcgcatgttcaagattgtagtctcctcttgagcaagttgtggaactacc 600
QY 601 aagcgtctgttttgacgaatgggttgcccaaatga 639
Db 601 aagcgtctgttttgacgaatgggttgcccaaatga 639

RESULT 3
AAV52155
ID AAV52155 standard; DNA; 28171 BP.
XX
AC AAV52155;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 273-289; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the

CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridize to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T; 0 other;

Query Match 99.5%; Score 635.8; DB 19; Length 28171;
Best Local Similarity 99.7%; Pred. No. 2.3e-190;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcaaaaggatttttagttctctcttgaggaccagaggagcagcaagaccagtgtt 60
Db 8013 atgtcaaaaggatttttagttctctcttgaggaccagaggagcagcaagaccagtgtt 8072
QY 61 ttgaggtctgtctaccaaatttttagaggaagaaagagtagaggtgttgacacccgtgaa 120
Db 8073 ttgaggtctgtctaccaaatttttagaggaagaaagagtagaggtgttgacacccgtgaa 8132
QY 121 cctggcggagcttgattggggagagattcgggaagttgttggttgatccaaagtcact 180
Db 8133 cctggcggagcttgattggggagagattcgggaagttgttggttgatccaaagtcact 8192
QY 181 cagatggtgctctaaacacagagctactctctctatttccagtcgcagacagcatttggtg 240
Db 8193 cagatggtgctctaaacacagagctactctctctatttccagtcgcagacagcatttggtg 8252
QY 241 gaaaaatttccagcccttgaaagctggcaagttggtgcatcgttctttatcgat 300
Db 8253 gaaaaatttccagcccttgaaagctggcaagttggtgcatcgttctttatcgat 8312
QY 301 agttctgtccctacaggaattggtcgtggttaagatatgaagccattgactggtc 360
Db 8313 agttctgtccctacaggaattggtcgtggttaagatatgaagccattgactggtc 8372
QY 361 aatcagtttgcagagatgggtcctcaaacccgatttgacactctatttgacatcgaggtg 420
Db 8373 aatcagtttgcagagatgggtcctcaaacccgatttgacactctatttgacatcgaggtg 8432
QY 421 gaagaagggtggtcgtatgtgtgcaatagtagcgcgaggttaatcgtttgattg 480
Db 8433 gaagaagggtggtcgtatgtgtgcaatagtagcgcgaggttaatcgtttgattg 8492
QY 481 gaaggttgacitgcataaaaaagttcgtcaaggctaaccttctctctggataaagag 540
Db 8493 gaaggttgacitgcataaaaaagttcgtcaaggctaaccttctctctggataaagag 8552
QY 541 ggaatcgcatgttcaagattgtagtctcctcttgagcaagttgtggaactacc 600
Db 8553 ggaatcgcatgttcaagattgtagtctcctcttgagcaagttgtggaactacc 8612
QY 601 aagcgtctgttttgacgaatgggttgcccaaatga 639
Db 8613 aagcgtctgttttgacgaatgggttgcccaaatga 8651

RESULT 4
AAZ96394
ID AAZ96394 standard; DNA; 1632 BP.
XX
AC AAZ96394;

PR 14-NOV-1997; 97US-0066009.

RESULT	7
AAV74612/c	
ID	AAV74612 standard; DNA; 4029 BP
XX	
XX	
AC	AAV74612;
XX	

DT XX 16-MAR-1999 (first entry)

DE XX Staphylococcus aureus contig SEQ ID #301.

XX

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;

KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;

KW toxic shock syndrome; ds.

XX

OS Staphylococcus aureus.

XX

XX

PH Key Location/Qualifiers

FT misc_feature 181..240

FT

FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT

FT misc_feature 1981..2040

FT

FT /tag= b

FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT

FT misc_feature 3781..3840

FT

FT /tag= c

FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT

PN EP786519-A2.

XX

XX 30-JUL-1997.

XX

XX 07-JAN-1997; 97EP-0100117.

XX

XX 05-JAN-1996; 96US-0009861.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

PI Rosen CA;

XX

XX WPI; 1997-374922/35.

DR

XX

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

XX

PS Claim 1; Page 1130-1133; 3271pp; English.

XX

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or

CC industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock

CC syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the

CC computer readable medium.

XX

XX Sequence 4029 BP; 1110 A; 793 C; 552 G; 1388 T; 186 other;

Query Match 18.7%; Score 119.4; DB 18; Length 4029;

Best Local Similarity 55.0%; Pred. No. 2.2e-27;

Matches 322; Conservative 0; Mismatches 251; Indels 12; Gaps 4;

QY 15 tttagctctcttgaggaccagagggagcaggaagaccaggtgttttagaggctctgt 74

DB 1257 TTTTATAACTTTTGGAGGCCAGAGAGGCTCTGGAAGAACACTGTAATTAATGAAGTTA 1198

QY 75 accaatttttagaggaagagtagaggtgttgacgagcccgtagaacctgcggaggtctt 134

DB 1197 CCAATAGATTA---GTAAAAGATTAATGATGTCATTATGACTAGAGAACCCAGGTGCTTCC 1141

QY 135 gattggggagagatcggggaagtgtattggatccaagtctactcagatcgatgctaa 194

DB 1140 TACTGGTGAAGNAATACGTAAATGTATAGAAAGG-----CAATGATATGGACATTTAG 1086

QY 195 aacgagagtaactctctatattgcccagtcgcagacagcatttgggtggaaaagtcttcc 254

DB 1085 AACTGAAGCAATGTTATTGCTGCATCTAGAAAGAGAACATCTTGTTATTAAGGTCTATACC 1026

QY 255 agccctgaagctggcgaagtgggtcatcatggtgttttatcgatagttctgttgctta 314

DB 1025 AGCTTTAAAGAAAGTAAGGTGTGTTGTCATGCTGCTATATCGATAGTTCATTAGCTTA 966

QY 315 tcagggtattggtcgtgcttagatattgaagccattgactggctcaatcagtttgcgac 374

DB 965 TCAAGGTATGCTAGAGGGATGCGGTTCAGAGAGTACAGACATTAACGAAATTTGCAAT 906

QY 375 agatgggctcaaacccgatttgacactctatttgacatcgaggtgggaagggctggc 434

DB 905 AATGGATTATATCCAGACTTGACGATTTATTTAAATGTTAGTGTGAAGTAGGTGCGGA 846

QY 435 tcgtatgctgcgaatagtagccgcgaggttaactgctgtgatttggaagggttggactt 494

DB 845 ACGTATTATT---AAAAATTCAGAGATCAAAATAGATTAGATCAAGAAAGATTTAAAGTT 789

QY 495 gcataaaaaagttcgtcaaggctacctttctctctggtgataaaggaggaaatcgcatgt 554

DB 788 -CAGGAAAAAGTAATTTGAAGGTTACCAAGAAATCATTCATTAATGAATCACACCGTTCAA 730

QY 555 caagattgatgctagctctcccttggagcaagttgttggaactac 599

DB 729 AAGCGTTAATGAGATCAACCTCTTGAAAAATGTTGTTGAAGACAC 685

RESULT 8

AAF28541

ID AAF28541 standard; DNA; 49617 BP.

XX AC AAF28541;

XX

XX DT 04-APR-2001 (first entry)

XX

XX Genomic fragment #28.

XX

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;

XX bronchopulmonary; endocarditis; meningitis; ss.

XX

XX Moraxella catarrhalis.

OS

XX WO200078968-A2.

PN

XX

XX PD 28-DEC-2000.

XX

XX PF 16-JUN-2000; 2000WO-US16649.

XX

XX PR 18-JUN-1999; 99US-0140121.

XX

XX PA (INCY-) INCYTE GENOMICS INC.

XX

XX PI Lagace RE, Patterson C, Berg KL;

XX

||||| |||| |||| ||| | ||||||||| |||| | || ||| ||||| ||| |||
Db 264382 ctg99gaagcgtataagcgatcgtttcoagattctactattgtatatcaaggaaattg 264443

Qy 326 gtctggcgttag 337
| | |||||
Db 264442 ctggaggcgttag 264453

RESULT 10
AAAX13181/c
ID ID AAAX13181 standard; DNA; 8102 BP.
XX
AC AAAX13181;
XX XX
XX DT 19-MAR-1999 (first entry)
XX XX
XX DE Enterococcus faecalis genome contig SEQ ID NO:244.
XX XX
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
OS OS
PN W09850555-A2.
XX XX
PD 12-NOV-1998.
XX XX
PF 04-MAY-1998; 98WO-US08985.
XX PF
PR 14-NOV-1997; 97US-0066009.
PR PR
PR 06-MAY-1997; 97US-0044031.
PR PR
PR 16-MAY-1997; 97US-00446655.
XX XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX XX
XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1210-1214; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAAX12938 to AAAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 8102 BP; 2343 A; 1656 C; 1404 G; 2696 T; 3 other;

Query Match 13.6%; Score 86.6; DB 20; Length 8102;
Best Local Similarity 57.2%; Pred. No. 7.4e-17;
Matches 155; Conservative 1; Mismatches 115; Indels 0; Gaps 0

QY 347 ccattgactggctccaacagtgttcgcacagatgggctcaaaccgccgatttgaccttatt 406
||||| | | ||| ||||||||| ||| | ||| ||||| ||| |||
Db 8101 CCATTGCTCAATTAATCGTTTGGCACAGAGGTGCTCACCAGATTTTACAATTTATT 8042

OY 407 ttgacatcgagdtgaaaagaaggcgtgctgtattgtgcgaatagtgaccqcgagtta 466

Db	8041	TAGACGTTGACTCAGATACCTGGCTTACGCGGATTCAAGAAATCGAACACAGSAAATTG	7982
Qy	467	atcgtttgattggaagggttgaccttgcatataaaaaagttcgtaaggctacattcttc	526
Db	7981	ATCGCTTAGATTACAGAAGGATTAGAAATTCATCAACGKGTTGCGTCATGAATATTGAAAT	7922
Qy	527	tctgataaagagggaatcgcatgtcaagattgatgctagtcctcctttggagcaag	586
Db	7921	TACCAAGAANAATCCACACGCTATCAAAAGATTGATGCACGGATGAGTTAGAACTTG	7862
Qy	587	tgtggaactaccaaggtgcttctgttga	617
Db	7861	TCGTAGAAGCAACGTATCAGCGGATTATTGA	7831
RESULT 11			
AAAF21608			
ID	AAAF21608 standard; DNA; 349980 BP.		
XX			
AC	AAAF21608;		
XX			
DT	13-MAR-2001 (first entry)		
XX			
OS	Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.		
DE			
XX			
KW	Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;		
KW	diagnosis; antigen; detection; infection; gene therapy; antibacterial;		
XX	ds.		
XX			
OS	Neisseria meningitidis.		
XX			
PN	WO2000066791-A1.		
XX			
PD	09-NOV-2000.		
XX			
PF	08-MAR-2000; 2000WO-US05928.		
XX			
PR	30-APR-1999; 98US-0132068.		
XX			
PR	08-OCT-1999; 99WO-US23573.		
XX			
PR	28-FEB-2000; 2000GB-0004695.		
XX			
XX	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX			
PI	Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;		
XX			
PI	Galottti C, Mora M, Rattli G, Scarselli M, Scarlato V, Rappuoli R;		
XX			
PI	Frazier CM, Grandi G;		
XX			
XX	WPI; 2000-647603/62.		
DR			
XX			
PT	Neisseria meningitidis B full length genome sequence and open reading		
XX	frames are used to detect, treat and prevent Neisserial infections -		
XX			
PS	Claim 7; Appendix A; 692pp; English.		
XX			
CC	The present invention describes the full length genome of		
XX			
CC	Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607		
XX			
CC	to AAF21613 represent fragments of the NMB genomic sequence, as the		
XX			
CC	sequence was too long to go in a record on its own it was split into 8		
XX			
CC	sequences which overlap each other at the beginning and end of each		
XX			
CC	sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at		
XX			
CC	the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at		
XX			
CC	the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the		
XX			
CC	Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to		
XX			
CC	AAAF21606 represent PCR primers which are used in the exemplification of		
XX			
CC	the present invention. The NMB genome and fragments from it have		
XX			
CC	antibacterial activity, and can be used in vaccines and gene therapy.		
XX			
CC	Neisseria nucleic acids, proteins and/or antibodies which binds to the		
XX			

KW M. genitalium; DNAA: DNA gyrase; origin of replication;
 XX megabase shotgun sequencing method; open reading frame; ORF; ss.
 OS Mycoplasma genitalium.
 FH Key Location/Qualifiers
 FT 8552..9184
 FT /tag= a
 FT /label= MG006
 FT /note= "Previously identified as MORF-20076, the
 FT encoded protein shows 27.59 percentage
 FT identity to thymidylate kinase (CDC8)
 FT from Saccharomyces cerevisiae"
 FT 11252..12040
 FT /tag= b
 FT /label= MG009
 FT /note= "Previously identified as MORF-20078, the
 FT encoded protein shows 35.43 percentage
 FT identity to the Bacillus subtilis hypothetical
 FT protein covered in accession number
 FT GB:D26185_102"
 FT 12069..12725
 FT /tag= c
 FT /label= MG010
 FT /note= "Previously identified as MORF-20079, the
 FT encoded protein shows 25.73 percentage
 FT identity to DNA primase (dnaE) from
 FT Clostridium acetobutylicum"
 FT complement (13570..14247)
 FT /tag= d
 FT /label= MG012
 FT /note= "Previously identified as MORF-20080, the
 FT encoded protein shows 31.50 percentage
 FT identity to the ribosomal protein S6
 FT modification protein (rimK) from Escherichia
 FT coli"
 FT complement (14396..15217)
 FT /tag= e
 FT /label= MG013
 FT /note= "Previously identified as MORF-19823, MORF-20080
 FT and MORF-20081, the encoded protein shows 33.04
 FT percentage identity to 5,10-methylene-tetra-
 FT hydrofolate dehydrogenase (fold) from E. coli"
 FT 17474..19243
 FT /tag= f
 FT /label= MG015
 FT /note= "Previously identified as MORF-20084, the
 FT encoded protein shows 32.23 percentage
 FT identity to transport ATP-binding protein
 FT (msbA) from E. coli"
 FT 26478..27344
 FT /tag= g
 FT /label= MG023
 FT /note= "Previously identified as MORF-20092, the
 FT encoded protein shows 45.96 percentage
 FT identity to fructose-bisphosphate aldolase
 FT (tsr) from B. subtilis"
 FT 27345..28448
 FT /tag= h
 FT /label= MG024
 FT /note= "Previously identified as MORF-19826 and
 FT MORF-20093, the encoded protein shows 46.84
 FT percentage identity to GTP-binding protein
 FT from E. coli"
 FT 36987..38978
 FT /tag= i
 FT /label= MG032
 FT /note= "Previously identified as MORF-20099, the
 FT encoded protein shows 26.82 percentage
 FT identity to ATP-dependent nuclease (addA)
 FT from B. subtilis"
 FT 39242..39904
 FT /tag= j

FT /label= MG033
 FT /note= "Previously identified as MORF-20100, the
 FT encoded protein shows 35.90 percentage
 FT identity to glycerol uptake facilitator
 FT (glpF) from B. subtilis"
 FT complement (39873..40514)
 FT /tag= k
 FT /label= MG034
 FT /note= "Previously identified as MORF-20101, the
 FT encoded protein shows 48.13 percentage
 FT identity to thymidylate kinase (tdk)
 FT from B. subtilis"
 FT 40543..41787
 FT /tag= l
 FT /label= MG035
 FT /note= "Previously identified as MORF-20102, the
 FT encoded protein shows 30.71 percentage
 FT identity to histidyl-tRNA synthetase (hiss)
 FT from Mycobacterium leprae"
 FT complement (44751..46277)
 FT /tag= m
 FT /label= MG038
 FT /note= "Previously identified as MORF-20105, the
 FT encoded protein shows 46.83 percentage
 FT identity to glycerol kinase (glpK)
 FT from E. coli"
 FT complement (46268..47422)
 FT /tag= n
 FT /label= MG039
 FT /note= "Previously identified as MORF-19831 and
 FT MORF-20106, the encoded protein shows 43.20
 FT percentage identity to glycerol-3-phosphate
 FT dehydrogenase (GUT2) from S. cerevisiae"
 FT 49377..49643
 FT /tag= o
 FT /label= MG041
 FT /note= "The encoded protein shows 48.86 percentage
 FT identity to phosphohistidinoprotein-hexose
 FT phosphotransferase (ptsH) from Mycoplasma
 FT capricolum"
 FT 50060..51520
 FT /tag= p
 FT /label= MG042
 FT /note= "Previously identified as MORF-19832 and
 FT MORF-20108, the encoded protein shows 41.92
 FT percentage identity to spermidine/
 FT putrescine transport ATP-binding protein
 FT (potA) from E. coli"
 FT 51525..52382
 FT /tag= q
 FT /label= MG043
 FT /note= "Previously identified as MORF-20110, the
 FT encoded protein shows 26.51 percentage
 FT identity to spermidine/putrescine transport
 FT system permease protein (potB) from E. coli"
 FT 52366..53220
 FT /tag= r
 FT /label= MG044
 FT /note= "Previously identified as MORF-20111, the
 FT encoded protein shows 29.45 percentage
 FT identity to spermidine/putrescine transport
 FT system permease protein C (potC) from E. coli"
 FT 54658..55605
 FT /tag= s
 FT /label= MG046
 FT /note= "Previously identified as MORF-20112, the
 FT encoded protein shows 36.60 percentage
 FT identity to sialoglycoprotease (gcp)
 FT from Pasteurella haemolytica"
 FT complement (56970..58310)
 FT /tag= t
 FT /label= MG048
 FT /note= "Previously identified as MORF-19834,

Db	8842	tgttttctctagtttgcataccaagagattaaacaaaataaggcgatgcatacagtaa	8901
Qy	354	ctggctcaatcagtttgcagacagatgggtcacaacccgatttgacactctattttgacat	413
Db	8902	acagattaatcatcatcgcttgaagaatatgatgccaaactttacctttattttggattg	8961
Qy	414	cgagttggaagaggttggtcgctgattgctgcgaatagtgacccgcgaggttaactcgttt	473
Db	8962	caattttaagaagcattacaagatgcacaaagcgtggaataatactctttgatga	9021
Qy	474	ggaatttgaaggttgagcttgcataaaaaagttctcaaggctacactttctctcttgga	533
Db	9022	atttataaggaagaatgatttgatcacgtctgtcttattattattaaagcttagttga	9081
Qy	534	tazagaggaaatcgcattgcgaattgatgcta	568
Db	9082	taaaaaaactgtttcttgattaatggtgataata	9116
RESULT 14			
AAH54919/c			
ID	AAH54919	standard; DNA; 3332 BP.	
XX	AAH54919;		
XX	03-SEP-2001	(first entry)	
XX			
DT			
XX			
DE		S. epidermidis genomic polynucleotide sequence SEQ ID NO:4283.	
XX			
KW		Staphylococcus epidermidis SR1 strain; infection; diagnosis;	
XX		vaccination; endocarditis; ds.	
XX			
OS		Staphylococcus epidermidis.	
XX			
PN	WO200134809-A2.		
XX			
PD	17-MAY-2001.		
XX			
PF	09-NOV-2000; 2000WO-US30782.		
XX			
PR	09-NOV-1999; 99US-0164258.		
XX			
PA	(GLAX) GLAXO GROUP LTD.		
XX	Kimmerly WJ;		
PI			
XX			
DR	WPI; 2001-316495/33.		
XX			
PT			
XX			
PT			
PT			
XX			
PS			
XX			
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides		
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.		
CC	(I) and (II) can have antibacterial activity and therefore can be used		
CC	in vaccination. The nucleic acids (I) may be used to produce the		
CC	S. epidermidis polypeptides (II) via the production of vectors		
CC	containing them which are used to produce hosts cells which express the		
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be		
CC	used to vaccinate subjects and to raise antibodies against the bacteria.		
CC	The polypeptides may also be used to assay for other inhibitors of their		
CC	activity and therefore identify compounds that may be used for the		
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to		
CC	AAH53090 represent specifically claimed S. epidermidis genomic DNA		
CC	polynucleotide sequences from the present invention. AAH55091 to		
CC	AAH55098 represent oligonucleotide sequences and primers which are used		
CC	in the exemplification of the present invention.		
CC	N.B. The present invention specifically claims all the polynucleotide		
CC	sequences given in the sequence listing of the present specification,		
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even		
CC	though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,		
CC	no sequences are present for SEQ ID NO:4455 to 4464.		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:16:17 ; Search time 37.91 Seconds
(without alignments)
414.232 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKFLVSLGPEGAGKTSV.....LEQVVTTKAVLFDGMLAK 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	100.0	212	21	AAB08637
2	1053	100.0	212	21	AAY81763
3	1047	99.4	212	19	AAY86073
4	528	50.1	212	20	AAY28785
5	475.5	45.2	214	22	AAB72200
6	475	45.1	205	21	AAY70146
7	354	33.6	206	20	AAY37013
8	329.5	31.3	209	20	AAY34868
9	306	29.1	210	20	AAY28787
10	305.5	29.0	213	20	AAY28786
11	305.5	29.0	213	22	AAB72201

12	300	28.5	210	20	AAY28792	Thymidylate kinase
13	277	26.3	210	20	AAY28791	Thymidylate kinase
14	200	19.0	188	20	AAY28789	Thymidylate kinase
15	198.5	18.9	191	19	AAW98749	H. pylori GPO 104
16	185.5	17.6	216	20	AAY28796	Yeast thymidylate
17	185.5	17.6	216	22	AAB72202	S. cerevisiae thym
18	160.5	15.2	224	21	AAG20878	Arabidopsis thalia
19	160.5	15.2	224	21	AAG50881	Arabidopsis thalia
20	160.5	15.2	263	21	AAG20877	Arabidopsis thalia
21	160.5	15.2	263	21	AAG50880	Arabidopsis thalia
22	130	12.3	394	22	AAG84942	Shrimp white spot
23	129	12.3	159	21	AAG20879	Arabidopsis thalia
24	129	12.3	159	21	AAG50882	Arabidopsis thalia
25	123.5	11.7	227	20	AAY28790	Thymidylate kinase
26	121.5	11.5	212	20	AAY28788	Human thymidylate
27	111.5	10.6	204	20	AAY28794	Thymidylate kinase
28	111.5	10.6	227	11	AA807563	Polypeptide encode
29	110.5	10.5	205	20	AAY28795	Thymidylate kinase
30	103	9.8	571	22	AAW16781	Peptide #3215 enco
31	103	9.8	571	22	AAW29266	Peptide #3303 enco
32	103	9.8	571	22	AAW04498	Peptide #3180 enco
33	102	9.7	54	22	AAW70417	A. actinomycetemco
34	102	9.7	203	22	AAW90585	C. glutamicum prote
35	102	9.7	203	22	AAW80129	Corynebacterium gl
36	101.5	9.6	210	20	AAY28793	Schizosaccharomyce
37	97.5	9.3	488	22	AAW31470	Amino acid sequenc
38	97	9.2	265	20	AAY03752	S. aureus Ferrichr
39	91	8.6	181	22	AAW90365	C. glutamicum prote
40	91	8.6	181	22	AAW80115	Corynebacterium gl
41	90.5	8.6	267	21	AAW82698	Aeropyrum pernix p
42	90	8.5	448	20	AAW34908	Chlamydia pneumoni
43	89.5	8.5	488	22	AAW31471	Amino acid sequenc
44	89.5	8.5	1045	20	AAW07482	Human chondrocyte-
45	89.5	8.5	1045	21	AAW91947	Human cytoskeleton

ALIGNMENTS

RESULT 1
AAB08637
ID AAB08637 standard; Protein; 212 AA.
XX
AC AAB08637;
DT 20-DEC-2000 (first entry)
XX
DE Amino acid sequence of a thymidylate kinase polypeptide.
XX
KW Thymidylate kinase; tdk polypeptide; bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO200050602-A1.
XX
PD 31-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04238.
XX
PR 26-FEB-1999; 99US-0259109.
XX
PR (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Burnham MKR, Zalacain M, Biswas S, Chalker AF, Ingraham KA;
PI Traini CM, Warren PV;
XX
DR WPI; 2000-533181/48.
XX
DR N-PSDB; AAA64398.
XX
PT Nucleic acids encoding thymidylate kinase family polypeptides derived
PT from Streptococcus pneumoniae, useful for screening for antibacterial
PT agents -
XX

PS Claim 1; Page 3; 4lpp; English.

CC The present sequence represents a polypeptide of the thymidylate kinase

CC family (tdk polypeptides), and is derived from Streptococcus pneumoniae.

CC tdk polynucleotides may be used to recombinantly produce the tdk

CC polypeptides either in vivo (e.g. as part of a genetic vaccination

CC procedure) or in vitro (e.g. as part of a fermentation culture).

CC The nucleic acids and proteins may be used to diagnose diseases

CC in which the tdk polypeptides are expressed, such as infection by

CC Streptococcus pneumoniae.

XX

SQ Sequence 212 AA;

Query Match 100.0%; Score 1053; DB 21; Length 212;

Best Local Similarity 100.0%; Pred. No. 1.1e-103;

Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKGLVLSLEGPEGAGKTSVLEALLPILKEKGVEVLTTRPFGVVLIGEKIREVILDPST 60

Db 1 mskglvlslegpegagktsvleallpilleekgvevlttrpfgvvligekevildpsht 60

Qy 61 QMDAKTELLLYASRRHLEKVLPALEAGKLVIMDRFIDSSVAYQVGRGLDTEADWL 120

Db 61 qmdakteullyasrrhlvekvlpaleagklvimdrfidssvayqvgfgrgldtealdwl 120

Qy 121 NOFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLEGLDHLKRVQGYLSLLDKE 180

Db 121 nqfatdglkpdltlyfdieeveglariaansdrevnrlidlegldhlhkvrgyyslldke 180

Qy 181 GNRIVKIDASPLEQVETTKAVLFDGMGLAK 212

Db 181 gnrvikidaspleqvettkavlfmgmlak 212

RESULT 2

AAy81763

ID AAY81763 standard; Protein; 212 AA.

XX

AC AAY81763;

XX

DT 02-JUN-2000 (first entry)

XX

DE Streptococcus pneumoniae protein sequence ID119.

XX

KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;

KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;

KW kidney disease; diabetes; immunosuppressive disorder; otitis media;

KW pneumococcal septicaemia; sinusitis; meningitis; therapy.

XX

OS Streptococcus pneumoniae.

XX

PN WO200006738-A2.

XX

PD 10-FEB-2000.

XX

PF 27-JUL-1999; 99WO-GB02452.

XX

PR 27-JUL-1998; 98GB-0016336.

PR 19-MAR-1999; 99US-0125329.

XX

PA (MICR-) MICROBIAL TECHNIQS LTD.

XX

PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

XX

DR WPI; 2000-195301/17.

DR N-PSDB; AAZ91859.

XX

PT Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections

XX

PS Claim 2; Page 61; 76pp; English.

XX

CC This sequence represents a Streptococcus pneumoniae protein of the

CC invention. The proteins (or their homologues, derivatives and/or

CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic

CC compositions comprising the proteins are useful as vaccines and also in

CC diagnostic assays. The sequences are useful for the detection or

CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested

CC with them. Agents capable of antagonising, inhibiting or interfering with

CC the function or expression of the protein or polypeptide are useful in

CC medical compositions in the treatment or prophylaxis of S. pneumoniae

CC infection. As the sequences can be used to treat S. pneumoniae infection,

CC they can be used to treat bacterial pneumonia, which has high rates in

CC young children, the elderly, and in patients with predisposing conditions

CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be

CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and

CC meningitis.

XX

SQ Sequence 212 AA;

Query Match 100.0%; Score 1053; DB 21; Length 212;

Best Local Similarity 100.0%; Pred. No. 1.1e-103;

Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKGLVLSLEGPEGAGKTSVLEALLPILKEKGVEVLTTRPFGVVLIGEKIREVILDPST 60

Db 1 mskglvlslegpegagktsvleallpilleekgvevlttrpfgvvligekevildpsht 60

Qy 61 QMDAKTELLLYASRRHLEKVLPALEAGKLVIMDRFIDSSVAYQVGRGLDTEADWL 120

Db 61 qmdakteullyasrrhlvekvlpaleagklvimdrfidssvayqvgfgrgldtealdwl 120

Qy 121 NOFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLEGLDHLKRVQGYLSLLDKE 180

Db 121 nqfatdglkpdltlyfdieeveglariaansdrevnrlidlegldhlhkvrgyyslldke 180

Qy 181 GNRIVKIDASPLEQVETTKAVLFDGMGLAK 212

Db 181 gnrvikidaspleqvettkavlfmgmlak 212

RESULT 3

AAy86073

ID AAY86073 standard; Protein; 212 AA.

XX

AC AAY86073;

XX

DT 10-APR-2000 (first entry)

XX

DE S. pneumoniae derived protein #282.

XX

KW Treatment; prevention; disease; diagnosis; gene therapy; screening;

KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.

XX

OS Streptococcus pneumoniae.

XX

PN WO9806734-A1.

XX

PD 19-FEB-1998.

XX

PF 15-AUG-1997; 97WO-US14436.

XX

PR 16-AUG-1996; 96US-0024022.

XX

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX

PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

PI Stodola RK;

XX

DR WPI; 1998-159452/14.

DR N-PSDB; AAZ96394.

XX

PT Streptococcus pneumoniae proteins and related DNA - useful for

XX

PR 13-FEB-1998; 98EP-0102546.

XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
XX PN WPI; 1999-508654/42.
XX PD
XX PF Producing kinases with increased activity on nucleoside and
XX PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
XX PR AZT, to active form
XX PS Claim 27; Page 71-72; 84pp; English.
XX PI The present sequence is the E.coli thymidylate kinase. It is a class II
CC (type II) kinase which generates cytotoxic nucleotide analogs from their
CC prodrugs, by phosphorylation, particularly the AIDS prodrug AZT.
CC Rational shuffling of domains from E.coli to the corresponding enzyme
CC from human, mouse, or yeast improves kinase activity. Increasing the
CC kinase activity results in higher concentrations of the active form of
CC the therapeutic analog (specifically AZT-triphosphate), and thus a
CC greater therapeutic effect, while reducing the concentration of the
CC toxic monophosphate intermediate. In E.coli TmpK the Arg15 in the P-loop
CC is absent while having basic residues in the LID region. Hence a P-loop
CC movement occurring upon AZT-MP binding should not have a detrimental
CC effect on catalysis. E.coli TmpK may have catalytic activity for
CC phosphorylation of AZT-monophosphate 300 times that of the wild-type
CC human enzyme. It has anticancer and antiviral activities.
XX SQ Sequence 210 AA;
Query Match 29.1%; Score 306; DB 20; Length 210;
Best Local Similarity 40.6%; Pred. No. 2.2e-24;
Matches 82; Conservative 36; Mismatches 74; Indels 10; Gaps 6;
Qy 3 KGFVLSLEGPGAGKTSVLEALLPILEKGV-EVLTTPREGVGLICEKIREVILDPSTQ 61
Db 2 kgkfivieglegagkssahqsvrvlhelgldvvtftregpgtqlaelrlhlikheteep 61
Qy 62 MDAKTELLLYTASRRQHLVEKVL-PALEAGKLVIMDRFIDSSVAYQVGFGRGLDIEADWL 120
Db 62 vtdkaellmlyaaariq-lvenvikpalmqgkwvgydrhdmsqayvgggrqldphfmitl 120
Qy 121 NOFATDGLKPDLYFDIEVEEGLARIANSREVNRLDLEGLDLHKVROGYLSLLDKE 180
Db 121 ketvlgnfepdltlyldpsvlgar--argrgeldrieqmdldffhrtraryielv-kd 177
Qy 181 GNRIKVIDASLPLEOVVETTKA 202
Db 178 npkavvina----eqsielvga 195
RESULT 10
AY28786
ID AAY28786 standard; protein; 213 AA.
XX AC AAY28786;
XX DT 02-NOV-1999 (first entry)
XX DE E.coli thymidylate kinase-1.
XX KW E.coli thymidylate kinase; E.coli TmpK; class II (type II) kinase; AIDS;
KW prodrug; cytotoxic nucleotide analog; AZT; 3'-azido-3'-deoxythymidine;
KW therapeutic analog; AZT-triphosphate; therapeutic effect; P-loop;
KW LID region; human thymidylate kinase; anticancer; antiviral.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT 12..17
XX FT /label= P-loop
XX FT /note= "unique additional basic residues"

FT Region 138..150
/label= LID_region
XX
XX PN WO9941404-A2.
XX PD 19-AUG-1999.
XX PF 12-FEB-1999; 99WO-EP00945.
XX PR 13-FEB-1998; 98EP-0102546.
XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
XX PN WPI; 1999-508654/42.
XX PD Producing kinases with increased activity on nucleoside and
XX PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
XX PR AZT, to active form
XX PS Claim 7; Page 70-71; 84pp; English.
XX PI The present sequence is the E.coli thymidylate kinase. It is a class II
CC (type II) kinase which generates cytotoxic nucleotide analogs from their
CC prodrugs, by phosphorylation, particularly the AIDS prodrug AZT.
CC Rational shuffling of domains from E.coli to the corresponding enzyme
CC from human, mouse, or yeast improves kinase activity. Increasing the
CC kinase activity results in higher concentrations of the active form of
CC the therapeutic analog (specifically AZT-triphosphate), and thus a
CC greater therapeutic effect, while reducing the concentration of the
CC toxic monophosphate intermediate. In E.coli TmpK the Arg15 in the P-loop
CC is absent while having basic residues in the LID region. Hence a P-loop
CC movement occurring upon AZT-MP binding should not have a detrimental
CC effect on catalysis. E.coli TmpK may have catalytic activity for
CC phosphorylation of AZT-monophosphate 300 times that of the wild-type
CC human enzyme. It has anticancer and antiviral activities.
XX SQ Sequence 213 AA;
Query Match 29.0%; Score 305.5; DB 20; Length 213;
Best Local Similarity 38.4%; Pred. No. 2.5e-24;
Matches 78; Conservative 37; Mismatches 79; Indels 9; Gaps 6;
Qy 7 VSLGEGAGKTSVLEALLPILEKGV-EVLTTPREGVGLICEKIREVILDPSTQMDA- 64
Db 6 ivieglegagktarnvvetleqgirdmvtftregpgtqlaelksirsvldiksvgdevi 65
Qy 65 --KTELLLYTASRRQHLVEKVL-PALEAGKLVIMDRFIDSSVAYQVGFGRGLDIEADWL 121
Db 66 tdkaelvlfyaarvq-lvetvikpalangtwigrdhldstqayvgggrgidqhmmlatlr 124
Qy 122 QFATDGLKPDLYFDIEVEEGLARIANSREVNRLDLEGLDLHKVROGYLSLLDKEG 181
Db 125 davgldfrpdtlyldtpevgklrararg--eidrieqesfdffnrtrarylelaaqd- 181
Qy 182 NRIVKIDASLPLEOVVETTKAVL 204
Db 182 ksihtidatqpleavmdairttv 204
RESULT 11
AAB72201
ID AAB72201 standard; protein; 213 AA.
XX AC AAB72201;
XX DT 04-MAY-2001 (first entry)
XX DE E. coli thymidylate kinase amino acid sequence.
XX KW Protein co-ordinate data; thymidylate kinase; TMK; binding pocket;

[illegible][illegible]

RESULT 12
AAAY28792
ID AAY28
XX
AC AAY28

RESULT 13
 AAY28791
 ID AAY28791 standard; protein; 210 AA.
 XX
 AC AAY28791;
 XX

DT 02-NOV-1999 (first entry)
 XX Thymidylate kinase-5.
 DE
 DE
 KW Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;
 KW viral replication; DNA chain termination; AZT activation pathway;
 KW AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;
 KW antiviral activity; therapeutic analog.
 XX
 OS Unidentified.
 XX
 XX WO9941404-A2.
 PN
 XX 19-AUG-1999.
 PD
 XX 12-FEB-1999; 99WO-EP00945.
 PF
 XX 13-FEB-1998; 98EP-0102546.
 PR
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA
 XX Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
 PI
 XX WPI; 1999-508654/42.
 DR
 XX Producing kinases with increased activity on nucleoside and
 PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
 PT AZT, to active form
 XX
 XX Claim 7; Page 76; 84pp; English.
 PS
 XX The present sequence is a thymidylate kinase. TmpK is involved in the
 XX activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication
 CC by DNA chain termination. TmpK which catalyses the second
 CC phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate
 CC (AZT-DP), is the rate limiting enzyme in the AZT activation pathway.
 CC Increasing the kinase activity of TmpK results in higher concentrations
 CC of the active form of the therapeutic analog especially AZT-triphosphate
 CC (AZT = 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,
 CC while reducing the concentration of the toxic monophosphate
 CC intermediate. This has anticancer and antiviral activities.
 XX
 SQ Sequence 210 AA;

Query Match 26.3%; Score 277; DB 20; Length 210;
 Best Local Similarity 36.4%; Pred. No. 2.6e-21;
 Matches 76; Conservative 37; Mismatches 74; Indels 22; Gaps 7;
 QY 1 MSKGLVSLGEGAGKTSVLEALLPILKEK--GVEVLTTPREGGVLGKIREVILDPSS 58
 Db 1 mkqgvfaiegvdgagktvllleafkqrfpgsflgftlfsrpggtplaeiralill--- 57
 QY 59 HTQMDAKTELLLYTASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYOGFGRGLDIEAID 118
 Db 58 heameplceaylfasarrctqrliqpqlqgqvivdrfwvssyayqglkkvgldvkv 117
 QY 119 WLNQFATDGLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDHLKK-----VRQG 172
 Db 118 klnadavgdmpdftfivdcdfetalnrmakrg--qdnld-----ntvkkqadfnmrcy 171
 QY 173 YLSLLDREGNRIKIDASLP---LEQVVE 198
 Db 172 yhsivdnk--rvflldgnqntgcicqqfie 198

RESULT 14
 AAY28789
 ID AAY28789 standard; protein; 188 AA.
 XX
 AC AAY28789;
 XX
 DT 02-NOV-1999 (first entry)

XX Thymidylate kinase-3.
 DE
 DE
 KW Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;
 KW viral replication; DNA chain termination; AZT activation pathway;
 KW AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;
 KW antiviral activity; therapeutic analog.
 XX
 OS Unidentified.
 XX
 XX WO9941404-A2.
 PN
 XX 19-AUG-1999.
 PD
 XX 12-FEB-1999; 99WO-EP00945.
 PF
 XX 13-FEB-1998; 98EP-0102546.
 PR
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA
 XX Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
 PI
 XX WPI; 1999-508654/42.
 DR
 XX Producing kinases with increased activity on nucleoside and
 PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
 PT AZT, to active form
 XX
 XX Claim 7; Page 74; 84pp; English.
 PS
 XX The present sequence is a thymidylate kinase. TmpK is involved in the
 XX activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication
 CC by DNA chain termination. TmpK which catalyses the second
 CC phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate
 CC (AZT-DP), is the rate limiting enzyme in the AZT activation pathway.
 CC Increasing the kinase activity of TmpK results in higher concentrations
 CC of the active form of the therapeutic analog especially AZT-triphosphate
 CC (AZT = 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,
 CC while reducing the concentration of the toxic monophosphate
 CC intermediate. This has anticancer and antiviral activities.
 XX
 SQ Sequence 188 AA;

Query Match 19.0%; Score 200; DB 20; Length 188;
 Best Local Similarity 31.0%; Pred. No. 3.3e-13;
 Matches 63; Conservative 36; Mismatches 68; Indels 36; Gaps 9;
 QY 1 MSKGLVSLGEGAGKTSVLEALLPILKEKGVLEVLTTPREGGVLGKIREVILDPSSHT 60
 Db 1 mvdnmfivfegldgsgkttgskll-----akkmdafwtvepsnslvgkllreil--sgkt 53
 QY 61 QMDAKTELLLYTASRRQH--LVEKVLPALEAGKLVIMDRFIDSSVAYOGFGRGLDIEAID 118
 Db 54 evdnktallfaadriehktlikeelkrd-----vvcdrylssiyasva-gvdenfik 108
 QY 119 WLNQFATDGLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDHLKKVRQGYLSLID 178
 Db 109 sinrya---lkpdivflilivdietaalkrvktkd-----ifekdkfkkvqdkyleae 158
 QY 179 KEGNRIVKIDASLPLEQVETTK 201
 Db 159 -eynfi-----vidttk 169

RESULT 15
 AAW98749
 ID AAW98749 standard; Protein; 191 AA.
 XX
 AC AAW98749;
 XX
 DT 31-MAR-1999 (first entry)
 XX

Search completed: February 15, 2002, 01:55:12
Job time: 2335 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:47:02 ; Search time 22.69 Seconds
(without alignments)
210.256 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKGFVLSLEGPEGAGKTSV.....LEQVVETTKAVLFDGMLAK 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	100.0	212	4	US-09-259-109-2
2	97	9.2	265	2	US-08-898-779-2
3	97	9.2	265	2	US-08-898-779-4
4	97	9.2	265	4	US-09-224-502-2
5	97	9.2	265	4	US-09-224-502-4
6	85	8.1	260	2	US-08-879-561-8
7	85	8.1	260	2	US-09-272-414-2
8	83.5	7.9	427	3	US-09-196-857-2
9	81.5	7.7	411	4	US-09-318-443-6
10	81	7.7	800	3	US-08-776-265-3
11	79.5	7.5	369	1	US-08-844-055-2
12	79.5	7.5	369	3	US-09-006-849-2
13	79.5	7.5	429	1	US-08-906-744A-2
14	79.5	7.5	429	3	US-09-093-134-2
15	76.5	7.3	313	4	US-09-396-651B-3
16	76.5	7.3	411	4	US-09-318-443-8
17	76	7.2	389	3	US-08-972-902-4
18	75.5	7.2	1544	4	US-09-413-814-46
19	75	7.1	503	3	US-08-911-853-7
20	75	7.1	503	4	US-09-479-409-7
21	74.5	7.1	292	2	US-08-928-284-2
22	74.5	7.1	547	2	US-08-467-822-35
23	74.5	7.1	547	4	US-08-432-697-35
24	74.5	7.1	547	4	US-08-466-248-35
25	74.5	7.1	876	1	US-08-717-515-4
26	74.5	7.1	1094	1	US-08-717-515-6
27	74.5	7.1	1276	1	US-08-717-515-8

Sequence 30, Appl
Sequence 20, Appl
Sequence 5, Appl
Sequence 12, Appl
Sequence 5, Appl
Sequence 3, Appl
Sequence 19, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 22, Appl
Sequence 22, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-259-109-2
; Sequence 2, Application US/09259109
; Patent No. 6270762
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Zalacaln, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Chalker, Alison F.
; APPLICANT: Ingraham, Karen A.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: tdk
; FILE REFERENCE: GM10201
; CURRENT APPLICATION NUMBER: US/09/259,109
; CURRENT FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-259-109-2

Query Match 100.0%; Score 1053; DB 4; Length 212;
Best Local Similarity 100.0%; Pred. No. 2e-106;
-Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSKGFVLSLEGPEGAGKTSVLEALLPILKEKGVLEVTREPGGVVLEKIREVILDPST 60
Db 1 MSKGFVLSLEGPEGAGKTSVLEALLPILKEKGVLEVTREPGGVVLEKIREVILDPST 60
QY 61 QMDAKTELLYIASRRHVLKVPALPALEAGKLVIMDRFIDSSVAYQGGRLDTEADWL 120
Db 61 QMDAKTELLYIASRRHVLKVPALPALEAGKLVIMDRFIDSSVAYQGGRLDTEADWL 120
QY 121 NQFATDGLKPDLTLYFFDIEVEEGLARIAANSRDVNRDLLEGLDHHKKVROGYSLLDKE 180
Db 121 NQFATDGLKPDLTLYFFDIEVEEGLARIAANSRDVNRDLLEGLDHHKKVROGYSLLDKE 180
QY 181 GNRIVKIDASLPLEQVVETTKAVLFDGMLAK 212
Db 181 GNRIVKIDASLPLEQVVETTKAVLFDGMLAK 212
RESULT 2
US-08-898-779-2
; Sequence 2, Application US/08898779
; Patent No. 5882891


```
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50444-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-224-502-2

Query Match          9.2%; Score 97; DB 4; Length 265;
Best Local Similarity 33.9%; Pred. No. 0.013;
Matches 40; Conservative 13; Mismatches 39; Indels 26; Gaps 6;

QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVLEVLTTRPFGVLLGKIREVILDPSTQMD 63
Db 29 GKVTSTIIGPNGCGKSTLLKALSRLAVKGEVF-----LDGENI-----HTQ-- 70

QY 64 AKTELLLYTASRRQHLVKVLPALRGLVIMDRFDISSVAYQGFR--GLDIEAIDW 119
Db 71 STKEIAKKIALLPQS--PEVADGLTVGELVSYGRFPHQ----KGFRLTAEDKKKEIDW 122

RESULT 5
US-09-224-502-4
; Sequence 4, Application US/09224502
; Patent No. 6264955
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; APPLICANT: Lonetto, Michael Arthur
; APPLICANT: Warren, Patrick Vernon
; TITLE OF INVENTION: NOVEL Ferrichrome transport
; TITLE OF INVENTION: ATP-Binding Protein
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,502
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/898,779
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50444-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; US-09-224-502-4

Query Match          9.2%; Score 97; DB 4; Length 265;
Best Local Similarity 33.9%; Pred. No. 0.013;
Matches 40; Conservative 13; Mismatches 39; Indels 26; Gaps 6;

QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVLEVLTTRPFGVLLGKIREVILDPSTQMD 63
Db 29 GKVTSTIIGPNGCGKSTLLKALSRLAVKGEVF-----LDGENI-----HTQ-- 70

QY 64 AKTELLLYTASRRQHLVKVLPALRGLVIMDRFDISSVAYQGFR--GLDIEAIDW 119
Db 71 STKEIAKKIALLPQS--PEVADGLTVGELVSYGRFPHQ----KGFRLTAEDKKKEIDW 122

RESULT 6
US-08-879-561-8
; Sequence 8, Application US/08879561
; Patent No. 5817482
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,561
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0325 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 181510
; US-08-879-561-8

Query Match          8.1%; Score 85; DB 2; Length 260;
Best Local Similarity 21.8%; Pred. No. 0.25;
Matches 54; Conservative 30; Mismatches 98; Indels 68; Gaps 8;

QY 7 VSLEGPPEGAGKTSVLEALLPILEKGV-----EVLTTREPQGVGLIGE 48
```

```
Db 24 ISIEGNTAAGKSTFVNILKQCEDEWVVPVPEVARWCNVQSTQDDEFEELTMSQKNG---GN 80
Qy 49 KIREVILDPSTQMDAKTELL-----YIASRRQHLVEKVLPALEAGKLVIMDRFIDSSV 103
Db 81 VLQMWYKPEPWSSTFTQYACLSIRAOQLASLNGKLKDAEKPLVFFERSVYSYRIIFASN 140
Qy 104 AYOGFGRGLDIEAT-----DWLNQFATDGLKPKDILTLYFDIEVEEGLARIANS 151
Db 141 LYES-----ECNNEETWIIQDWHDMNMQFGQSLDGIIVLOATPFTCLHRIYLRG 193
Qy 152 DREVRNLDLEGLD-----LHKKVRQGY-----LSLLDKEGRIIVKIDASLPLE 194
Db 194 RNEEQGIPLEYLEKHLHYKHESWLLHRTLTNTFVYLQEVPIITLD-----VNEDFKKYE 247
Qy 195 QVETTKAVL 204
Db 248 SLVEKVEFL 257

RESULT 7
US-09-727-414-2
; Sequence 2, Application US/09272414A
; Patent No. 6238885
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: Histidine Kinase
; FILE REFERENCE: GM10202
; CURRENT APPLICATION NUMBER: US/09/272,414A
; CURRENT FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-727-414-2
```

```
Query Match 8.1%; Score 85; DB 4; Length 583;
Best Local Similarity 25.6%; Pred. No. 0.86; Mismatches 23; Indels 40; Gaps 7;
Matches 46; Conservative 23;

Qy 35 VLTTRPGVGLGKIREVILDPSTQMDAKTELLYIASR-RQHLVEKVLPALEAGKLV 93
Db 269 VLGINESRQIILSNKMANDIM--NIDEDAKAFLLRQIEDTFKSKQTEMRDLENNAREFV 326
Qy 94 IMDRFIDSSVAYQFGRGLDIEADWLNQFATDGLKPD-----LTYFDI 138
Db 327 VTTSYIDK--TEQGGKSGVVTVTRDMTNEHNDQMKKDFIANVSHELRTPISLQGYTES 384
Qy 139 -----EVERGLARIANSR-----EVNRDLEGLDLHKVRQGYLSLLDK 179
Db 385 IVDGIVTEPDEIKESLAWVLDESRLNRLNELLNVARMDAEGLSVNKEV-OPTAALLDK 443
```

```
RESULT 8
US-09-196-857-2
; Sequence 2, Application US/09196857A
; Patent No. 6100069
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Brown, James R.
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Throup, John P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Zhong, YiYi
; APPLICANT: Debouck, Christine
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Wang, Min
; APPLICANT: Warren, Richard L.
; APPLICANT: Schilling, Lisa K.
```

```
; TITLE OF INVENTION: No. 6100069el tig
; FILE REFERENCE: GM10118
; CURRENT APPLICATION NUMBER: US/09/196,857A
; CURRENT FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: 60/082,418
; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-196-857-2
```

```
Query Match 7.9%; Score 83.5; DB 3; Length 427;
Best Local Similarity 24.1%; Pred. No. 0.78; Mismatches 36; Indels 73; Gaps 14;
Matches 61; Conservative 36;

Qy 20 VLEALLP-----ILEEKGEV-----LTTREPG-----GVLIGE-KIREV 53
Db 67 VMNALLPNAYEAAYKEAGLEVVAQPKIDVTSMEKGQDWIAAEVVTKEPVKLDYKKNLEV 126
Qy 54 ILDPSTQMDAKTELLYIASRRQHLVEKVL--PALEAGKLVIMDRFIDS--SVAYQG-- 107
Db 127 SVDEKEVTDADVE--ERIERERNLTVELIKEAAENGDTWID-FVGSIDGVFEFDGCK 183
Qy 108 ---FGRGLDI-----EADWLNQFATDGLKPKD-----LTYFDIE 139
Db 184 GENFSLGIGSQGFIPGFEDQLVGHSGAGTVDIVITFPEDYOADLAGKEAFVTTIHEVK 243
Qy 140 VEEGLA---RIAANSREVRNRLDLEGLDLHKVRQGYLSLLD---KEGNRIIVKIDASLP 193
Db 244 AKEVPALDDELAKDIDEVETL----ADLKEKYKELAAAKEAYKDAVEGAADTAVEN 299
Qy 194 EQVETTKAVLFD 206
Db 300 AEIVELPEEMIHE 312
```

```
RESULT 9
US-09-318-443-6
; Sequence 6, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-318-443-6
```

```
Query Match 7.7%; Score 81.5; DB 4; Length 411;
Best Local Similarity 23.4%; Pred. No. 1.2; Mismatches 77; Indels 75; Gaps 12;
Matches 57; Conservative 35;

Qy 3 KGFVLSLEGPEGAGKT-----SVLEAL-LPILKEKGEVLTRE----- 40
Db 74 KGRDVIAQSQSGTGCTATFSSVLQCLDIQVRETQALITAPRELAVQIQKGLLAGDYM 133
Qy 41 -----PGGVLIKEKIREVILDPSTQMDAKTELLYIASRRQHLVEKVLPALE---BAG 90
Db 134 NVQCHACTGGTNGVEDIRK--LDYQG-HVVAGTGPCRVFDMIRRRSLRTRAIKMLVLEAD 190
```

QY 91 KLV-----IMD--REFSSVAYQFGRLDIEADLWLNQFATDGLKPDLTLYFDIEVE 141
Db 191 EMLNKGFEQIYDVRPLSATVVLISATLPHLEILNTKFMFDPI----- 237
QY 142 EGLARIAANSREVNRLDLEGLDLHKVKVROGYLSLLDKEGNRIVKIDASLPLEQVETTK 201
Db 238 ----RILVKRD---ELTLEG-----IKOFFVAVEREW-----KFDTLCDLYDTLITQ 279
QY 202 AVL F 205
Db 280 AVIF 283
RESULT 10
US-08-776-265-3
; Sequence 3, Application US/08776265
; Patent No. 6001631
; GENERAL INFORMATION:
; APPLICANT: BLANCHE, Francis
; APPLICANT: CAMERON, Beatrice
; APPLICANT: CROUZET, Joel
; APPLICANT: FAMECHON, Alain
; APPLICANT: FERRERO, Lucia
; TITLE OF INVENTION: No. 6001631el Topoisomerase IV, Corresponding
; TITLE OF INVENTION: Nucleotide Sequences and Uses Thereof
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I. Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,265
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 03806.0394-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4444
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-776-265-3
Query Match 7.7%; Score 81; DB 3; Length 800;
Best Local Similarity 28.1%; Pred. No. 3.8;
Matches 48; Conservative 29; Mismatches 62; Indels 32; Gaps 9;
QY 50 IREVILDPSTQMDA---KTELLLYIASRQHLVEKVLPALEAGKLVIMDRFID---SSV 103
Db 343 IROIISYLNHQIEVAVNRKTFELDNAEKRMHIVEGLIKALS-----ILDKVIELRSK 397
QY 104 AYQFGRLDIEADLWLNQFATDGLKPDLTLY----FDIEVEEG-----LARIA 148
Db 398 NKRAKENL-IEVEFTEEGAEALVM--LQLYRLTNTDIVALEGEHKEALIKQLRHIL 454
QY 149 ANSDREVNRLDLEGLDLHKVKVROGYLSLLDKEGNRIVKID--ASLPLEQV 197

Db 455 DNHALLNVIKEELNEIKKKFKSERLSLIEAEIEEI-KIDKEVWVPSEVI 504
RESULT 11
US-08-844-055-2
; Sequence 2, Application US/08844055
; Patent No. 5747313
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5747313el Compounds
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,055
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 369 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-844-055-2
Query Match 7.5%; Score 79.5; DB 1; Length 369;
Best Local Similarity 20.1%; Pred. No. 1.7;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
QY 6 LVSLEGPE--GAGKTSVLEALLPI-----LEEKGVELTTRPGGVLTGE---K 49
Db 165 LNTLGNPESRAAYRQALIDVLTPLKETLSKDSQRLEENPLRVLDSEKEDKAVENAPS 224
QY 50 IREVILDPSTQMDAKTELL--YIASRR-----QHLVEKVLPALEAGKLVIM- 95
Db 225 ILDFLDEESQTHFDVAVSQMLENLGVYDIIDTNMVRGLDYNNHTIFETIEGNDLTVCA 284
QY 96 -DRFIDSSVAY-----QGFGRLDIEADLWLNQFATDGLKPDLTLYFDIEVEEG LARIA 148
Db 285 GGRY-DGLVAYFGGPGTAGFGGLGVERL----- 312
QY 149 ANSDREVNRLDLEGLDLHKVKVROGYLSLLDKEGNRIVKIDASLPLEQVETTKAVLFDG 207
Db 313 -----LLILEKQG-----VALPIENALDVYIAVLGDG 339
RESULT 12
US-09-006-849-2
; Sequence 2, Application US/09006849

Patent No. 6071731
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: NO. 6071731e1 Compounds
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,849
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-849-2

Query Match 7.5%; Score 79.5; DB 3; Length 369;
Best Local Similarity 20.1%; Pred. No. 1.7;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
Qy 6 LVSLGGPE--GAGKTSVLEALLPI-----LEEKGEVLTTRPGGVGLIGE---K 49
Db 165 LNTLGNPESRAAYRQALIDYLTPLKETSQSRRLRNPLRVLDSEKEDKVAVENAPS 224
Qy 50 IREVILDPSTQMDAKTELLL-----YIASRR-----OHLVEKVLPALEAGKLVIM- 95
Db 225 ILDFDEESQTHFDVAVSQMLNGLVDYIIDTNMVRGLDYNNHTIFEITEIGNDLTVCA 284
Qy 96 -DRFIDSSVAY-----QGFGRLDIEADWLNOFATDGLKPDLTLYFDIEVEEGLARIA 148
Db 285 GGRY-DGLVAYFGGPETAGFGGLGVERL-----LILEKOG-----VALPIENALDVYIAVLGDG 312
Qy 149 ANSDREVNRLDLEGLDLHKVQRQGYLSLLDKENRIVKIDASLPLEQVVTTKAVLFDG 207
Db 313 -----LILEKOG-----VALPIENALDVYIAVLGDG 339

RESULT 13
US-08-906-744A-2
Sequence 2, Application US/08906744A
Patent No. 5795758
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Greenwood, Rebecca
APPLICANT: Lawlor, Elizabeth

TITLE OF INVENTION: NOVEL HISS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,744A
FILING DATE: 06-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-906-744A-2

Query Match 7.5%; Score 79.5; DB 1; Length 429;
Best Local Similarity 20.1%; Pred. No. 2.1;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
Qy 6 LVSLGGPE--GAGKTSVLEALLPI-----LEEKGEVLTTRPGGVGLIGE---K 49
Db 165 LNTLGNPESRAAYRQALIDYLTPLKETSQSRRLRNPLRVLDSEKEDKVAVENAPS 224
Qy 50 IREVILDPSTQMDAKTELLL-----YIASRR-----OHLVEKVLPALEAGKLVIM- 95
Db 225 ILDFDEESQTHFDVAVSQMLNGLVDYIIDTNMVRGLDYNNHTIFEITEIGNDLTVCA 284
Qy 96 -DRFIDSSVAY-----QGFGRLDIEADWLNOFATDGLKPDLTLYFDIEVEEGLARIA 148
Db 285 GGRY-DGLVAYFGGPETAGFGGLGVERL-----LILEKOG-----VALPIENALDVYIAVLGDG 312
Qy 149 ANSDREVNRLDLEGLDLHKVQRQGYLSLLDKENRIVKIDASLPLEQVVTTKAVLFDG 207
Db 313 -----LILEKOG-----VALPIENALDVYIAVLGDG 339

RESULT 14
US-09-093-134-2
Sequence 2, Application US/09093134
Patent No. 6040162
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Greenwood, Rebecca
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: NOVEL HISS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/093,134
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/906,744
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-093-134-2

Query Match 7.5%; Score 79.5; DB 3; Length 429;
Best Local Similarity 20.1%; Pred. No. 2.1;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
QY 6 LVSLEGPE--GAGKTSVLEALLPI-----LEEKGEVLTTRPGGVLTGE---K 49
DB 165 LNTLGNPESRAYQALIDYLTPLKETLSKDSQRRLLENPLRVLDSKEKDKAVENAPS 224
QY 50 IREVLPDSHTQMDAKTELL-----YIASRR-----QHLVEKVLPALEAGKLVIM- 95
DB 225 ILDFLDEESQTHFDVAVSQMLENLGVYDIIDTNMYRGLDYNHHTFEITEIEGNDLTVCA 284
QY 96 -DRFIDSSVAY-----QGFGRLDIEAIDMLNFAFDGLKPOLTYFDIEVEEGLARIA 148
DB 285 GGRY-DGLVAYFGGPETAGFGGLGVERL-----LULEKQ-----VALPIENALDVYIAVLGDG 339
QY 149 ANSDREVNRLDLEGLDLHKVRQGLSLLDKEGNIIVKIDASLPLEQVETTKAVLFDG 207
DB 313 -----LULEKQ-----VALPIENALDVYIAVLGDG 339

RESULT 15
US-09-396-651B-3
Sequence 3, Application US/09396651B
Patent No. 6225076
GENERAL INFORMATION:
APPLICANT: Darst, Seth A
APPLICANT: Zhang, Gongyi
APPLICANT: Campbell, Elizabeth
APPLICANT: Minakin, Leonid
APPLICANT: Severinov, Konstantin
TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS

TITLE OF INVENTION: OF USE THEREOF
FILE REFERENCE: 600-1-258
CURRENT APPLICATION NUMBER: US/09/396,651B
CURRENT FILING DATE: 1999-09-15
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 313
TYPE: PRT
ORGANISM: Thermus aquaticus
US-09-396-651B-3

Query Match 7.3%; Score 76.5; DB 4; Length 313;
Best Local Similarity 24.2%; Pred. No. 2.8;
Matches 47; Conservative 35; Mismatches 75; Indels 37; Gaps 10;
QY 32 GVEVLTTRPG-GVLIGEKEIREVILD--PSHTQMDAKTELLY----IASRQHLVEKVL 84
DB 21 GEFVLEPLERGFVTLGNPLRILLSSIPGTAVTSVYIEDVLHEFSTIPGVKEDVVEIL 80
QY 85 PALEAGKLVIMDRFID----SSVAVQGFGRGLDIEAIDMLNQFATDGLKPLD----- 132
DB 81 NLKE---LVV--REFDPWRWTLILRAEG-PKEVRAVDFTPSADVEIMNPDLHIATLEEG 134
QY 133 -TLYFDIEVEEGLARIAANSRDREVNRLDLEGLD-LHKVRQ-----GYLSLLDKE 180
DB 135 GKLYMEVVRVDRGVGVPAERHGIKDRINAIPVDAIFSPVRRVAFQVEDTRLQRTDLDKL 194
QY 181 GNRIVKIDASLPLE 194
DB 195 TLRIWTGDSVTPL 208

Search completed: February 15, 2002, 01:55:56
Job time: 534 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:50:02 ; Search time 33.21 Seconds
(without alignments)
486.269 Million cell updates/sec

Title: US-09-749-972-2
Perfect score: 1053
Sequence: 1 MSKGLVSLGPEGAGKTSV.....LEQVVTTKAVLFDGMGLAK 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	659	62.6	211	2 F86674	thymidylate kinase
2	551	52.3	210	2 B83655	thymidylate kinase
3	528	50.1	212	2 S66058	thymidylate kinase
4	404	38.4	210	2 E83275	thymidylate kinase
5	392.5	37.3	195	2 H70383	thymidylate kinase
6	387.5	36.8	206	2 D75560	thymidylate kinase
7	373	35.4	1089	2 T36663	protein kinase, tr
8	368.5	35.0	197	2 E72294	thymidylate kinase
9	361	34.3	203	2 E71546	probable thymidyla
10	361	34.3	203	2 G81700	thymidylate kinase
11	357.5	34.0	206	2 H81932	probable dtmP kina
12	356	33.8	206	2 H81173	thymidylate kinase
13	329.5	31.3	206	2 A86525	thymidylate kinase
14	329.5	31.3	206	2 F72098	thymidylate kinase
15	314	29.8	212	2 A82128	thymidylate kinase
16	307.5	29.2	205	2 A75165	thymidylate kinase
17	306	29.1	237	2 A64153	dtmP kinase (EC 2.
18	305.5	29.0	213	2 G64853	dtmP kinase (EC 2.
19	304.5	28.9	213	2 H85672	thymidylate kinase
20	300	28.5	210	2 F64200	thymidylate kinase
21	294.5	28.0	196	2 E69257	thymidylate kinase
22	294	27.9	205	2 A71177	thymidylate kinase
23	289.5	27.5	203	2 G71674	probable thymidyla
24	280	26.6	210	2 S73474	thymidylate kinase
25	262	24.9	212	2 A84971	dtmP kinase (EC 2.
26	262	24.9	233	2 H82944	thymidylate kinase
27	238.5	22.6	199	2 H84343	thymidylate kinase
28	204.5	19.4	217	2 F82788	thymidylate kinase
29	201	19.1	98	2 S48605	hypothetical prote

30	200	19.0	188	2 F64336	dtmP kinase (EC 2.
31	198.5	18.9	191	2 B64704	thymidylate kinase
32	195.5	18.6	191	2 B71814	thymidylate kinase
33	185.5	17.6	216	1 K1BY78	dtmP kinase (EC 2.
34	174	16.5	208	2 H71333	probable thymidyla
35	172	16.3	194	2 F69102	thymidylate kinase
36	170.5	16.2	192	2 F81347	thymidylate kinase
37	169	16.0	208	2 E72514	probable dtmP kina
38	154	14.6	236	2 S76121	hypothetical prote
39	145	13.8	257	2 H70198	thymidylate kinase
40	133.5	12.7	252	2 F65084	hypothetical prote
41	131.5	12.5	230	2 H65084	hypothetical prote
42	129	12.3	188	2 T52029	dtmP kinase (EC 2.
43	125.5	11.9	210	2 T41553	thymidylate kinase
44	123	11.7	211	2 S26845	dtmP kinase (EC 2.
45	115.5	11.0	218	2 T24244	hypothetical prote

ALIGNMENTS

RESULT 1

F86674

thymidylate kinase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: F86674

R;Bolotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.;
Genome Res. in press, 2001

A;Title: The complete genome sequence of the lactic acid bacterium.

A;Reference number: A86625

A;Accession: F86674

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-211 <STO>

A;Cross-references: GB:AE005176; NID:gl2723271; PIDN:AAK04496.1; GSPDB:GN00146

A;Experimental source: strain IL1403

C;Genetics:

A;Gene: yeaB

C;Superfamily: dtmP kinase

Query Match 62.6%; Score 659; DB 2; Length 211;
Best Local Similarity 61.7%; Pred. No. 2e-42;
Matches 124; Conservative 39; Mismatches 38; Indels 0; Gaps 0;

QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVLEVLTREPQGVLTGKIREVILDPSTQMD 63

Db 3 GILSLGPDGAGKTTVLKEILPEIQMKREIVPTREPQGVVABEIRQIILDPKNTDID 62

QY 64 AKTELLLYIASRRHLEKVLPALEAGKLVIMDRFIDSSVAVQGGRLGIDEAIDMLAQF 123

Db 63 SKTEMLFAAARRLHMOEKMDLPALQAGKVIIVDRFIDSSVAVQGGRLGVEVDMLNYF 122

QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRVNRLDLBGLDHLKRVGYSLLDKEGSR 183

Db 123 ATDGLKPDLTLYFDVDIDVALERIMKNRADEVNRLDLERAEMHRKVRGYLEIVVKEPER 182

QY 184 IVKIDASLPLEQVVTETTKAVL 204

Db 183 FVKIDASQPLEKVVADTILSVL 203

RESULT 2

B83655

thymidylate kinase tmk [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C;Accession: B83655

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.,

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans

A;Reference number: A83650; MUID:20263314

A:Accession: B93655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AP001507; GB:BA000004; NID:gi0172612; PIDN:BA03761.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: tmk
C:Superfamily: dtMP kinase

Query Match 52.3%; Score 551; DB 2; Length 210;
Best Local Similarity 54.8%; Pred. No. 2.4e-34;
Matches 108; Conservative 38; Mismatches 51; Indels 0; Gaps 0;

QY 1 MSKGLVSLGEGAGKTSVLEALLPILKEKGVEVLTTRPGGVILGKIREVILDPST 60
DB 1 MTKGCFIVGEGAGKTSALDAIEMLENGLSVVRTPGPGIPAIQIRSIILDVHT 60

QY 61 QMDATELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWL 120
DB 61 RMDPRTEALLYAARRQHLVEKVLPALEAGHVLCDFRIDSLSLAYQGYARGIGFEDILAI 120

QY 121 NQFATDGLKPDLTLYFDIEVEEGLARIAANSRVNRDLLEGDLHKKVRQGYLSLLDKE 180
DB 121 NEFAIEGRYPDLTLLFRYDPDVGSLRIHQDSREQNRDLQDEALTFOHKVKEGYRIVETY 180

QY 181 GNRIKVIDASLPLEQV 197
DB 181 PERVVEIDANQSFQV 197

RESULT 3
S66058
thymidylate kinase tmk - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jan-2001
C:Accession: S66058; D69724
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
A:Reference number: S65967; MUID:96051385
A:Accession: S66058
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <OGA>
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05264.1; PID:g467418
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: D69724
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <KUN>
A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11804.1; PID:g2632295
A:Experimental source: strain 168
C:Genetics:
A:Gene: tmk
C:Superfamily: dtMP kinase

C:Keywords: nucleotide binding; P-loop
F:10-17/Region: nucleotide-binding motif A (P-loop)

Query Match 50.1%; Score 528; DB 2; Length 212;
Best Local Similarity 51.7%; Pred. No. 1.3e-32;
Matches 106; Conservative 37; Mismatches 62; Indels 0; Gaps 0;

QY 4 GFLVSLGEGAGKTSVLEALLPILKEKGVEVLTTRPGGVILGKIREVILDPSTQMD 63
DB 3 GLFITFEGEGAGKTTVLEIKNLTABGLQVMATREPGGIDIAEQIREVILNENILMD 62

QY 64 AKTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNOF 123
DB 63 PKTEALLYAARRQHLVEKVLPALEAGKLVLCDFRIDSLSLAYQGYARGGLDIEVLSINEF 122

QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRVNRDLLEGDLHKKVRQGYLSLLDKEGNR 183
DB 123 AIGDMMPHVTVYFSDPEGLKRIYANGSREKNRLDLEKDPHTKVQEGYQELMKRFFPR 182

QY 184 IVKIDASLPLEQVETTKRAVLFDGM 208
DB 183 PHSVDAGQSKDLVQDVLKVIDEAL 207

RESULT 4
E83275
thymidylate kinase PA2962 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83275
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: AB2950; MUID:20437337
A:Accession: E83275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AE004722; GB:AE004091; NID:g9949054; PIDN:AAG06350.1; GSPDB:G
A:Experimental source: strain PA01
C:Genetics:
A:Gene: tmk; PA2962
C:Superfamily: dtMP kinase

Query Match 38.4%; Score 404; DB 2; Length 210;
Best Local Similarity 47.7%; Pred. No. 2.4e-23;
Matches 92; Conservative 25; Mismatches 74; Indels 2; Gaps 1;

QY 4 GFLVSLGEGAGKTSVLEALLPILKEKGVEVLTTRPGGVILGKIREVILDPSTQMD 63
DB 3 GLFITFEGEGAGKSTNDRYLAERLRERIEVQLTRPGGTPPLAERIPELLAPSDEPMA 62

QY 64 AKTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNOF 123
DB 63 ADTELLMFAARQAHLGAVIRPALARGAVLCDFRDTATYAYQGGGRLPEARIALESF 122

QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRVNRDLLEGDLHKKVRQGYLSLLDKEGNR 183
DB 123 VQGLRDLPLTLVFDLPVETGLARAARG--RLDRFEQEDRRRFEAVRQTYLQRAAQAPER 180

QY 184 IVKIDASLPLEQV 196
DB 181 YQVLDAGLPLAEV 193

RESULT 5
H70383
thymidylate kinase - Aquifex aeolicus
C:Species: Aquifex aeolicus

Query Match 35.0%; Score 368.5; DB 2; Length 197;
Best Local Similarity 42.6%; Pred. No. 1e-20;
Matches 81; Conservative 39; Mismatches 59; Indels 11; Gaps 4;

Qy 7 VSLEGGAGKTSVLEALLPILKEKGVVLTTRPGGVILGKIREVILDPSTQMDAKT 66
Db 3 ITFEGIDSGSTQIQLLAQYLEKRGKVKILKREPGGTETGEKIRKILLE---EETPKA 59

Qy 67 ELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNQFATD 126
Db 60 ELFLFLASRNL-LYTEIKQYLSEGVAVLLDRYTDSSVAYQGFGRNLGKIEIVEELNDFATD 118

Qy 127 GLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDLHKVKRGYLSLLDKEGNRIVK 186
Db 119 GLIPDLTYIDVDVETALKR-----KGELNR--FEKREFLEVRREGYLVLAAREHPRIW 171

Qy 187 IDASLPLEQV 196
Db 172 LQKRSIEEI 181

RESULT 9
E71546
probable thymidylate kinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71546
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:99000809
A:Accession: E71546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <ARN>
A:Cross-references: GB:AE001292; GB:AE001273; NID:g3328586; PIDN:AAC67780.1; PID:g332859
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: tdk
C:Superfamily: dTMP kinase

Query Match 34.3%; Score 361; DB 2; Length 203;
Best Local Similarity 40.9%; Pred. No. 3.9e-20;
Matches 83; Conservative 40; Mismatches 60; Indels 20; Gaps 6;

Qy 5 FLVLEGGAGKTSVLEALLPILKEKGVVLTTRPGGVILGKIREVILDPSTQMDA 64
Db 2 FIV-VEGEGAGKTQFTQALSKRLIEGREGREIVTTREPGGCSLGDSVRLGLDPEQ-KISP 59

Qy 65 KTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNQFA 124
Db 60 YAEILLFLAARAHQIHEKILPALESGKTVICDRFHDSTIVYQGIAGGLG-----EAFV 112

Qy 125 TD-----GLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDLHKVKRGYLS 175
Db 113 TDLCVRVVGDEPFLPDITFLDLPEKGLLR--KTRQNLDRFEQKPTSFHRAAREGIS 170

Qy 176 LLDKEGNRIVKIDASLPLEQVVE 198
Db 171 LAERAPDRYKVLDALEPTEASVD 193

RESULT 10
G81700
thymidylate kinase TC0460 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: G81700
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3
A:Reference number: A81500; MUID:20150255
A:Accession: G81700
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <TET>
A:Cross-references: GB:AE002314; GB:AE002160; NID:g7190495; PIDN:AAF39311.1; PID:g71
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0460
C:Superfamily: dTMP kinase

Query Match 34.3%; Score 361; DB 2; Length 203;
Best Local Similarity 39.9%; Pred. No. 3.9e-20;
Matches 81; Conservative 40; Mismatches 62; Indels 20; Gaps 5;

Qy 5 FLVLEGGAGKTSVLEALLPILKEKGVVLTTRPGGVILGKIREVILDPSTQMDA 64
Db 2 FIV-VEGEGAGKTQFTQALSKRLMEEGKEVLTREPSSALGELQRLDLVDVTQ-EISS 59

Qy 65 KTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNQFA 124
Db 60 YAEILLFLAARAHQIHEKILPALESGKTVICDRFHDSTIVYQGIAGGLG-----EAFV 112

Qy 125 TD-----GLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDLHKVKRGYLS 175
Db 113 TDLCVRVVGDEPFLPDITFLDLPEKGLLR--KTRQNLDRFEQKPTSFHRAAREGIS 170

Qy 176 LLDKEGNRIVKIDASLPLEQVVE 198
Db 171 LAERAPDRYKVLDALEPTEASVD 193

RESULT 11
H81932
probable dTMP kinase (EC 2.7.4.9) NMA0869 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81932
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 2249
A:Reference number: A81775; MUID:2022556
A:Accession: H81932
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84149.1; PID:g73
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: tmk; NMA0869
C:Superfamily: dTMP kinase
C:Keywords: phosphotransferase

Query Match 34.0%; Score 357.5; DB 2; Length 206;
Best Local Similarity 39.9%; Pred. No. 7.2e-20;
Matches 79; Conservative 41; Mismatches 75; Indels 3; Gaps 2;

Qy 7 VSLEGGAGKTSVLEALLPILKEKGVVLTTRPGGVILGKIREVILDPSTQMDAKT 66
Db 6 ITLDIGAGKSTNLAVIKAWFERRGPLVLTREPPTPGVGEALREILNP-ETKAGURA 64

Qy 67 ELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDWLNQFATD 126
Db 65 ETLMWFAARMOHIEDVILPALSDGCHVVSDFRTATAYQGGGMPSEDIIEHWWQV 124

Qy 127 GLKPDLTLYFDIEVEGLARIAANSREVNRLDLEGLDLHKVKRGYLSLLDKEGNRIVK 186
Db 125 GLRPDLTLLDVPLEVSMARI--QOTREKDRFEQEQAFMRVRSVRLNRAAACPERYAV 182

Search completed: February 15, 2002, 01:56:50
Job time: 408 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:55:57 ; Search time 24.29 Seconds
(without alignments)
320.006 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKGLVSLGPEGAGKTSV.....LEQVETTKAVLFDGMGLAK 212

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	659	62.6	211	1 KTHY_LACLA	Q9ci94 lactococcus
2	551	52.3	210	1 KTHY_BACHD	Q9kg19 bacillus ha
3	528	50.1	212	1 KTHY_BACSU	P37537 bacillus su
4	404	38.4	210	1 KTHY_PSEAE	Q9hzn8 pseudomonas
5	392.5	37.3	195	1 KTHY_AQUAE	O67099 aquifex aeo
6	387.5	36.8	206	1 KTHY_DEIRA	Q9ry40 deinococcus
7	368.5	35.0	197	1 KTHY_THEMEA	Q9x013 thermotoga
8	361	34.3	203	1 KTHY_CHLMO	Q9pkk5 chlamydia m
9	361	34.3	203	1 KTHY_CHLTR	O84191 chlamydia t
10	357.5	34.0	206	1 KTHY_NEIMA	Q9jve7 neisseria m
11	356	33.8	206	1 KTHY_NEIMB	Q9k0d9 neisseria m
12	337	32.0	208	1 KTHY_CAUCR	Q9rq19 caulobacter
13	329.5	31.3	206	1 KTHY_CHLPN	Q9z8f5 chlamydia p
14	314	29.8	209	1 KTHY_PASMO	Q9cke9 pasteurella
15	314	29.8	212	1 KTHY_VIBCH	Q9kq12 vibrio chol
16	307.5	29.2	205	1 KTHY_PYRAB	Q9v1e9 pyrococcus
17	306	29.1	210	1 KTHY_HABIN	P44719 haemophilus
18	305.5	29.0	213	1 KTHY_ECOLI	P37345 escherichia
19	305	29.0	212	1 KTHY_YERPE	O99169 yersinia pe
20	300	28.5	210	1 KTHY_MYCGE	P47252 mycoplasma
21	294.5	28.0	196	1 KTHY_ARCFU	O30175 archaeoglob
22	294	27.9	205	1 KTHY_PYRHO	O59366 pyrococcus
23	289.5	27.5	203	1 KTHY_RICPR	Q9zcn9 rickettsia
24	280	26.6	210	1 KTHY_MYCPN	P75106 mycoplasma
25	262	24.9	212	1 KTHY_BUCAI	P57434 buchnera ap
26	262	24.9	230	1 KTHY_UREPA	Q9prc5 ureaplasma
27	238.5	22.6	199	1 KTHY_HALN1	Q9hmv4 halobacteri
28	200	19.0	188	1 KTHY_METJA	O57741 methanococ
29	198.5	18.9	191	1 KTHY_HELPY	O26009 helicobacte
30	198.5	18.9	208	1 KTHY_XYLFA	Q9pf57 xyella fas
31	195.5	18.6	191	1 KTHY_HELPJ	Q9zjes helicobacte
32	188	17.9	189	1 KTHI_SULSO	Q9uxg7 sulfolobus
33	185.5	17.6	216	1 KTHY_YEAST	P00572 saccharomyc

34	172	16.3	194	1	KTHY_METTH	O27793 methanobact
35	170.5	16.2	192	1	KTHY_CAMJE	Q9ppf3 campylobact
36	169	16.0	208	1	KTHY_AERPE	Q9ya48 aeropyrum p
37	154	14.6	212	1	KTHY_SYNV3	Q55593 synechocyst
38	146	13.9	193	1	KTHY_THEAC	Q9hlz2 thermoplasm
39	146	13.9	213	1	KTH2_SULSO	P58157 sulfolobus
40	133.5	12.7	252	1	YGHR_ECOLI	Q46842 escherichia
41	131.5	12.5	230	1	YGHR_ECOLI	Q46844 escherichia
42	125.5	11.9	210	1	KTHY_SCHPO	P36590 schizosacch
43	123.5	11.7	227	1	KTHY_MOUSE	P97930 mus musculu
44	121.5	11.5	212	1	KTHY_HUMAN	P23919 homo sapien
45	115.5	11.0	218	1	KTHY_CAEEL	Q22018 caenorhabdi

ALIGNMENTS

```

RESULT 1
KTHY_LACLA
ID KTHY_LACLA STANDARD; PRT; 211 AA.
AC Q9CIG4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus;
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RL lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006276; AAK04496.1; -
DR InterPro; IPR000062; Thymidylate_kin.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 211 AA; 23999 MW; 1CC1F2C5ED9B2A55 CRC64;

```

Query Match 62.6%; Score 659; DB 1; Length 211;
Best Local Similarity 61.7%; Pred. No. 2.6e-40;
Matches 124; Conservative 39; Mismatches 38; Indels 0; Gaps 0;

Oy 4 GFLVSLGPEGAGKTSVLEALLPILKEGVEVLITREPGGVILGKIREVILDPSTQMD 63

Db 3 GILISLEGPDGAGKTTVLKEILPEIQKMKREIVTPREPGGVRVAEIRQILDPKNTDID 62

Oy 64 AKTELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSVAVGQFGRGRLDTEADWLNOF 123

[illegible][illegible]

```

Db 121 NEFALEGYRYPDLTLLFRVDDPGLVSRIHRDQSGREONRLDQEALTFHQVKKEGYERIVETY 180

Qy 181 GNRIYKIDASLPLEQVV 197
      !!!!!!!
Db 181 PERVVEIDANQSFDQVV 197

RESULT 3
KTHY_BACSU
ID KTHY_BACSU STANDARD; PRT; 212 AA.
AC P37537;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1] CSEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP FUNCTION.
RA Gilles A.M., Barzu O.;
RL Unpublished observations (JUN-1996).
CC -1- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-MONOPHOSPHATE = ADP +
CC THYMIDINE 5'-DIPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D26185; BAA05264.1; -.
DR EMBL; Z99104; CAB11804.1; -.
DR Subtilist; BG10092; tmk.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide blosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17
SQ SEQUENCE 212 AA; 23876 MW; 0E4519621128FA16 CRC64;
```

Query Match	50.1%	Score 528;	DB 1;	Length 212;
Best Local Similarity	51.7%;	Pred. No. 5.2e-31;		
Matches 106;	Conservative 37;	Mismatches 62;	Indels 0;	Gaps 0;
QY	4	GFLVSLGPGGAGKTVSLVLEALLPILEKGEVLEVTTRPPGGVLICEKIREVILDRSHQMD	63	
Db		: : : : : : : :		
	3	GLFTFPGPGAGKTVLQEIKNLTAEGLQVMTAREPGGIDIAEQIREVILNENNLMD	62	
QY	64	AKTELLYIASRRGHLVEKVLPALEACKVLITMDRFDISSVAYQGFGRGLDTEADWLNOF	123	
Db		: : :		
	63	PKEALLYAAARRGHLVEKVKPALEQGFIVLCDFDISPLAYQGYARGLGIDEVLSINEF	122	
QY	124	ATDGLKPDLTLYFDIEVEEGLIARTAAANSDRVNRDLLEGDLHKKVRQGYLSLLDKGNR	183	
Db		: : : : :		
	123	AIGDMPHVAVYFSDIEPEGLKRTYANGSKREKRLDLEKLDFFHPKVOEGYELMKRPER	182	

RESULT 5
KTHY_AQUAE
ID KTHY_AQUAE STANDARD: PRT; 195 AA.

AC	67059	STANDARD	FAL	155 RA.
AT	15-JUL-1999	(Rel. 38, Created)		
CT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	THYMIDYLATE KINASE (EC 2.7.4.9) (DMP KINASE)			
DN	TKM OR AQ169			
GC	Aquifex aequificus			
OS	Bacteria; Aquificales; Aquificaceae; Aquifex			

NCBI_TaxID=63363;
[1]
SEQUENCE FROM N.A.
STRAIN=VF5;

RX	MEDLINE=98196666; PubMed=9537320;
RA	Decker G., Warren P.V., Gaasterland T., Young
RA	Graham D.E., Overbeek R., Snead M.A., Keller M
RA	Feldman R.A., Short J.M., Olson G.J., Swanson
RT	"The complete genome of the hyperthermophilic
RT	aeolicus.";
RL	Nature 392:353-358(1998).
CC	-!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM
CC	NOVO AND SALVAGE PATHWAYS OF DTP SYNTHESIS
CC	-!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-MONO
CC	THYMIDINE 5'-DIPHOSPHATE.
CC	-!- SIMILARITY: BELONGS TO THE THYMIDYLATE KIN
CC	-----

This SWISS-PROT entry is copyright. It is produced between the Swiss Institute of Bioinformatics

the European Bioinformatics Institute. There use by non-profit institutions as long as

CC modified and this statement is not removed. W

CC
CC
CC
CC
CC
CC
CC
CC
CC
EMBL: AE000716; AAC07063.1;
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin.1;
DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1.
KW Transfrase; Kinase; Nucleotide biosynthesi;

Complete proteome.	ATP (POTENTIAL)
Complete proteome.	14
NP BIND	7
FT	7
KW	7

SEQUENCE 195 AA; 22370 MW; F3C2A94D4930956

Country	Year	Value
China	2007	1.0
China	2008	1.0
China	2009	1.0
China	2010	1.0
China	2011	1.0
China	2012	1.0
China	2013	1.0
China	2014	1.0
China	2015	1.0
China	2016	1.0
China	2017	1.0
China	2018	1.0
China	2019	1.0
China	2020	1.0
China	2021	1.0
China	2022	1.0
China	2023	1.0
China	2024	1.0
China	2025	1.0
China	2026	1.0
China	2027	1.0
China	2028	1.0
China	2029	1.0
China	2030	1.0
China	2031	1.0
China	2032	1.0
China	2033	1.0
China	2034	1.0
China	2035	1.0
China	2036	1.0
China	2037	1.0
China	2038	1.0
China	2039	1.0
China	2040	1.0
China	2041	1.0
China	2042	1.0
China	2043	1.0
China	2044	1.0
China	2045	1.0
China	2046	1.0
China	2047	1.0
China	2048	1.0
China	2049	1.0
China	2050	1.0
China	2051	1.0
China	2052	1.0
China	2053	1.0
China	2054	1.0
China	2055	1.0
China	2056	1.0
China	2057	1.0
China	2058	1.0
China	2059	1.0
China	2060	1.0
China	2061	1.0
China	2062	1.0
China	2063	1.0
China	2064	1.0
China	2065	1.0
China	2066	1.0
China	2067	1.0
China	2068	1.0
China	2069	1.0
China	2070	1.0
China	2071	1.0
China	2072	1.0
China	2073	1.0
China	2074	1.0
China	2075	1.0
China	2076	1.0
China	2077	1.0
China	2078	1.0
China	2079	1.0
China	2080	1.0
China	2081	1.0
China	2082	1.0
China	2083	1.0
China	2084	1.0
China	2085	1.0
China	2086	1.0
China	2087	1.0
China	2088	1.0
China	2089	1.0
China	2090	1.0
China	2091	1.0
China	2092	1.0
China	2093	1.0
China	2094	1.0
China	2095	1.0
China	2096	1.0
China	2097	1.0
China	2098	1.0
China	2099	1.0
China	2100	1.0

Query Match	37.3%	Score	392.3	DB
Best Local Similarity	45.5%	Pred. No.	1.9e-21	

Matches 87; Conservative 42; Mismatches 5

```

Db      2  LIAFEGIDSGKTTQAKKLYEYLKQGYFVSLYRPGGTVKQGVFLREILLT---EELDER 58
QV     66  TELLVYASRRHOHVEKVPALFAGKLVIMDRFIDSSVAYOGGGRGIDFAIDWNOFAT 125

```

59	TELLIEFASPSKIEEKIITDINKOKWIIIDREVI	STIAVCVCVKCIIDVEEIKNIINEEAT	118
0b			

04	1 36	DGI KDOI MI VEDTQVBECCI ABTAANENDEWNBFI DI BCI DI HXKVBQCVI S DVECNMTIV	195
08	37	***** EMBACONNENFEBR111 DEBNDAVA VIJMA AKYI ITOI QOIGNSDSDTETI ZANNENETITU	111

[illegible]

DD 119 KGVKPDIIIELEDDIPVDIALKRL-----KEANK--FENKEFLEAVKNGFDELAKKEEN-VV 177

186 KIDASLPLEQV 196
QY
||| | : |

Db 171 VIDASGEEEV 181

RESULT 6

KTHY DEIRA

KTHY DEIRA

```
ID KTHY_DEIRA STANDARD; PRT; 206 AA.
AC Q9RY40;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK OR DR011.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001873; AAF09698.1; -
DR HSSP; P37345; 5TMP.
DR TIGR; DR0111; -
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 11 18 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 22430 MW; E506BDFDC9F63D CRC64;

Query Match 36.8%; Score 387.5; DB 1; Length 206;
Best Local Similarity 43.4%; Pred. No. 4.7e-21;
Matches 85; Conservative 36; Mismatches 72; Indels 3; Gaps 2;

QY 1 MSKGLFVSLGEGAGKTSVLEALLPILKEKGVETLTREPGGVILGKIREVILDPSPHT 60
DB 1 MSQGLFITLEGEGAGKTTQLARLEARLAAGHAVTVTRPGGTPGTRVREVLDPA-V 59
QY 61 OMDAKTELLYTSRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRLDIEAIDWL 120
DB 60 EIEPGEFLYKSASRQLVREVRPALENGETVLCDRADSSAYQAGAGRLSLPLRQI 119
QY 121 NOFATDGLKPDILTYFDIEVEGLARIANSREYNRLDLEGLDLHKVRQGYLSLLDKE 180
DB 120 TAEVTGGTLPGLTLLDLPALGLQRAARRG--QPRLEQADLTTHRRVRQGFLLAHAE 177
QY 181 GNRIVKIDASPLEQV 196
DB 178 PQRFVLDAATPEDEL 193
```

RESULT 7
KTHY_THEMA

6--4

```
ID KTHY_THEMA STANDARD; PRT; 197 AA.
AC Q9X0I3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK OR TM1099.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS88 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 395:323-329(1999).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001769; AAD36175.1; -
DR TIGR; TM1099; -
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 7 14 ATP (POTENTIAL).
SQ SEQUENCE 197 AA; 22849 MW; 302D2EF3BAB6658E CRC64;

Query Match 35.0%; Score 368.5; DB 1; Length 197;
Best Local Similarity 42.6%; Pred. No. 9.9e-20;
Matches 81; Conservative 39; Mismatches 59; Indels 11; Gaps 4;

QY 7 VSLEPGEGAGKTSVLEALLPILKEKGVETLTREPGGVILGKIREVILDPSPHTOMDAKT 66
DB 3 ITFEGIDSGKSTQTLQAQVLEKRGKVKILKREPGETGKIRKILLE---EETPKA 59
QY 67 ELLYIAGRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRLDIEAIDWLNOFATD 126
DB 60 ELFLASRNLL-LVTEIKQYVSEGVAVLLDRYDSSVAYQFGFNLGHEIVEELNDFATD 118
QY 127 GLKPDILTYFDIEVEGLARIANSREYNRLDLEGLDLHKVRQGYLSLLDKEGNRIVK 186
DB 119 GLIPDLTYFDVDTALKR-----KGELNR--FEKREFLERVREGVLVLAHEHPERIV 171
QY 187 IDASPLEQV 196
DB 172 LDGKRISIEI 181
```

RESULT 8

KTHY_CHLMU

ID KTHY_CHLMU

AC Q9PKK5;

STANDARD; PRT; 203 AA.


```

RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parkey D.S., Blair E., Cittiene H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE002421; AAF41088.1; -.
DR TIGR; NMB0670; -.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 22986 MW; 2D356EFE373E8F5E CRC64;
Query Match 33.8%; Score 356; DB 1; Length 206;
Best Local Similarity 40.8%; Pred. No. 8e-19;
Matches 82; Conservative 38; Mismatches 77; Indels 4; Gaps 3;
QY 7 VSLEPGCAGKTSVLKALLPILEKGVEVLTTREPGGVLTGKIREVILDPSHTMDAKT 66
Db ::::|::::| : : : | : | | | | : | : | : | : | : | : | : | :
Db 6 ITLDGDGACGKSTNLAVIKAWFERRGPLVLTREPGGTPVGCEALREILLNP-ETKAGLA 64
QY 67 ELLLYIARQHLYEKVLPALAEAGKLVMDFRIDSSVAYOGFGRGDLDEAIDWLNOFATD 126
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 65 ETLMWFARQHQHTEEVLPALSDGIHVVSFRDTATFAYOGGGMPSEDIIELEHWVQG 124
QY 127 GLKPDLTYFDIEVEEGARIANAANSDEVNRNLDLEGDLHKVKRQGYLSLLDGKNRIK 186
Db ||||| | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 125 GLKPDLTLLLDVPLEVSMARI--GQTRKDREFEQADFMFVRGVYLDRAAACPERYAV 182
QY 187 IDASLPLEQVVETTKAVLFDDG 207
Db || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 183 IDSNNELDEVNRNSTEKKVL-DG 202
RESULT 12
KTHY_CAUCR STANDARD; PRT; 208 AA.
AC Q9RQJ9;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK OR CC1824.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxId=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
```

RA Ohta N., Newton A.;
 RT "dnac encodes the delta prime subunit of DNA polymerase III.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / Cbl5;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Feidlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.K.R., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Shapiro L., Venter J.C., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE
 CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
 CC ADP + THYMIDINE 5'-DIPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF099189; AAF06831.1; -;
 DR EMBL: AE005856; AAK23799.1; -;
 DR TIGR: CC1824; -;
 DR InterPro: IPR00062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1
 DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1.
 KW Transferrase; Kinase; Nucleotide biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 11 18 ATP (POTENTIAL).
 SQ SEQUENCE 208 AA; 22107 MW; 7BD15DFD05F42D02 CRC64;
 Query Match 32.08; Score 337; DB 1; Length 208;
 Best Local Similarity 37.68; Pred. No. 1.8e-17;
 Matches 79; Conservative 34; Mismatches 95; Indels 2; Gaps 1;
 QY 1 MSKGLVLEGEPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVGVLGKIREVILDPST 60
 DB 1 MTQGFIFTEGEGAGKTSVLEALLPILKEKGVLEVTTRPGGVGVLGKIREVILDPST 60
 QY 61 QMDAKTELLYIASRRQHLVKEKVPALPALEAGKLVIMDRFIDSSVAYQGFGRGLDTEAIDWL 120
 DB 61 RWSPTVTELLMYAARRDHIERVIRGLARGAVVLCDFEADSTRAYQAGGDAPASLTAAL 120
 QY 121 NQFATDGLKPDLTLYFDIEVEGLARIAANSREVRNLDLEGLDLHKVQGYLSLDKE 180
 DB 121 EHVGLGTVPTVLTLLDPAEVLQRAEARG--AARFESKGLAFHERLRAGYLEIARRE 178
 QY 181 GNRIVKIDASLPLEQVQVETTKAVLFDGMGL 210
 DB 179 PDRCVIDAADAEALDAVTAISDVVVQRLGL 208
 RESULT 13
 KTHY_CHLPN
 ID KTHY_CHLPN STANDARD; PRT; 206 AA.
 AC Q928R5; Q9J094;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTPP KINASE).
 GN TWK OR CPN0273 OR CP0486.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99266606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA Linhe O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE
 CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
 CC ADP + THYMIDINE 5'-DIPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001612; AAD18422.1; -;
 DR EMBL: AE002210; AAF38316.1; -;
 DR EMBL: AP002346; BAA98483.1; -;
 DR TIGR: CP0486; -;
 DR InterPro: IPR00062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1.
 DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1.
 KW Transferrase; Kinase; Nucleotide biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 7 14 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 22852 MW; 42B0E8AA5262768 CRC64;
 Query Match 31.38; Score 329.5; DB 1; Length 206;
 Best Local Similarity 35.38; Pred. No. 6.1e-17;
 Matches 76; Conservative 46; Mismatches 72; Indels 21; Gaps 5;
 QY 5 FLVLEGEPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVGVLGKIREVILDPSTOMDA 64
 DB 2 FIV-IEGEGSGKSLAKALGDQLVADQRKVLITREPCCGLIGLRDLILPPHLELSR 60
 QY 65 KTELLYIASRRQHLVKEKVPALPALEAGKLVIMDRFIDSSVAYQGFGRGLDTEAIDWLNOFA 124
 DB 61 CCELFELGSRQAQHQEVIIPALRDGYIVICERFHDSTIVYOGIAEGLAD-----FV 113
 QY 125 TD-----GLKPDLTLYFDIEVEGLARIAANSREVRNLDLEGLDLHKVQGYLS 175
 DB 114 ADLCSKVVGPTTFFLPNPFVLLDIPADIGLQR--KHKQKVFQKFKPLSYHNRIREGFLS 171

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:55:17 ; Search time 55.75 Seconds
(without alignments)
556.228 Million cell updates/sec

Title: US-09-749-972-2
Perfect score: 1053
Sequence: 1 MSKGLVSLGEGAGKTSV.....LEQVVETTKAVLFDGMGLAK 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTEMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	670.5	63.7	211	2 Q9A190	Q9A190 streptococ
2	653	62.0	211	2 Q9A0D5	Q9A0D5 lactococ
3	475	45.1	205	2 Q9WCL1	Q9WCL1 staphyloc
4	373	35.4	1089	2 Q9X908	Q9X908 streptomyce
5	365	34.7	226	2 Q9EZH7	Q9EZH7 xanthomonas
6	233	22.1	165	2 Q9R2M7	Q9R2M7 versinia en
7	232	22.0	165	2 Q9R2M9	Q9R2M9 versinia en
8	232	22.0	165	2 Q9R2M8	Q9R2M8 versinia en
9	231	21.9	165	2 Q9R1B4	Q9R1B4 versinia ps
10	230	21.8	165	2 Q9R1E6	Q9R1E6 versinia ps
11	230	21.8	165	2 Q9R2Y0	Q9R2Y0 versinia ps
12	201	19.1	98	2 Q49001	Q49001 mycoplasma
13	174	16.5	208	2 Q83373	Q83373 treponema p
14	145	13.8	257	2 Q51733	Q51733 borrelia bu
15	144	13.7	309	3 Q9C2R7	Q9C2R7 neurospora
16	141	13.4	388	12 Q9DQD0	Q9DQD0 white spot
17	134.5	12.8	212	11 Q9D7I3	Q9D7I3 mus musculus
18	129	12.3	188	10 Q81650	Q81650 arabidopsis
19	120.5	11.4	212	4 Q9BUX4	Q9BUX4 homo sapien

20	107	10.2	80	2 Q9A29	Q9A29 vibrio mari
21	106.5	10.1	210	2 Q9CCJ3	Q9CCJ3 mycobacteri
22	103.5	9.8	205	2 Q9W52	Q9W52 staphylococ
23	103	9.8	333	4 Q9Y804	Q9Y804 aeropyrum p
24	103	9.8	903	4 Q9BWL4	Q9BWL4 homo sapien
25	102	9.7	310	2 Q9PE05	Q9PE05 xylella fas
26	101	9.6	168	3 Q9HFB8	Q9HFB8 thanatephor
27	101	9.6	337	1 Q9YDH3	Q9YDH3 aeropyrum p
28	101	9.6	405	1 Q28871	Q28871 archaeoglob
29	100	9.5	248	1 Q9YCF6	Q9YCF6 aeropyrum p
30	100	9.5	314	2 Q9KEN4	Q9KEN4 bacillus ha
31	99.5	9.4	250	1 Q58948	Q58948 pyrococcus
32	99.5	9.4	288	2 Q9KF34	Q9KF34 bacillus ha
33	99.5	9.4	328	1 Q9V0X0	Q9V0X0 pyrococcus
34	98.5	9.4	1482	3 Q74637	Q74637 penicillium
35	98	9.3	239	2 Q31427	Q31427 bacillus su
36	98	9.3	608	1 Q9UZN6	Q9UZN6 pyrococcus
37	97.5	9.3	214	2 Q05891	Q05891 mycobacteri
38	97.5	9.3	1499	3 Q9Y839	Q9Y839 mycosphaere
39	97	9.2	265	2 Q9X665	Q9X665 staphylococ
40	97	9.2	605	1 Q58850	Q58850 pyrococcus
41	96.5	9.2	171	5 Q9V8B5	Q9V8B5 drosophila
42	96.5	9.2	558	2 Q9JR75	Q9JR75 neisseria m
43	95.5	9.1	503	2 Q9CLB3	Q9CLB3 pasteurella
44	95.5	9.1	1619	3 Q13407	Q13407 magnaporthe
45	95	9.0	374	1 Q46492	Q46492 acidianus a

ALIGNMENTS

RESULT 1
Q9A190 PRELIMINARY; PRT; 211 AA.
ID AC Q9A190;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE THYMIDYLATE KINASE.
GN TMK OR SPY0399.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lal H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: A5006502; AAK33434.1;
KW Kinase; Complete proteome.
SQ SEQUENCE 211 AA; 23402 MW; 3866FBI39A188752 CRC64;

Query Match	63.7%	Score	670.5;	DB	2;	Length	211;
Best Local Similarity	67.7%	Pred. No.	4.8e-39;				
Matches	134;	Conservative	30;	Mismatches	33;	Indels	1;
Gaps	1;						
QY	1	MSKGLVSLGEGAGKTSVLEALLPILEEK-GVEVLTTREPPGGVLTGEKIREVILDP	59				
Db	1	MITGKLTIVGPGDAGKTTVLEQLIPLLKQKVAQDILTTREPPGGVLTGEKIREVILDP	60				
QY	60	TQMDAKTELLYIASRRQHLVKEVLPALPALEAGKLVIMDRFDISSVAYQGFGRGLDIEADW	119				
Db	61	TAMPDKTELLYIAARRQHLVKEVLPALPALEAGKLVIMDRFDISSVAYQGFGRGLDIEADW	120				
QY	120	LNQFATDGLKPDLTLYFDIEVEEGLARIANSREVRNLDLEGLDLHKVKRQGLSLDK	179				

Db 121 LNEFATDGLPDLTLYFDVPSEIGLARINANOQREVNRLDLETIEHQRVKGYLALAKE 180
QY 180 EGNRIKVIDASLPLEQV 197
Db 181 HPKRIVTIDATKPLKEV 198

RESULT 2
Q9AQD5 ID Q9AQD5 PRELIMINARY; PRT; 211 AA.
AC Q9AQD5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TSCORF1
GN TSCORF1984.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGI363;
RX MEDLINE=20244636; PubMed=10784052;
RA O'Connell-Motherway M., van Sinderen D., Morel-Deville F.,
Fitzgerald G.F., Ehrlich S.D., Morel P.;
RT "Six putative two-component regulatory systems isolated from
Lactococcus lactis subsp. cremoris MGI363.";
RL Microbiology 146:935-947(2000).
DR EMBL; AF178425; AAG53731.1; -.
SQ SEQUENCE 211 AA; 23973 MW; 94ADB5F3F6985EFA CRC64;

Query Match 62.0%; Score 653; DB 2; Length 211;
Best Local Similarity 62.2%; Pred. No. 7.8e-38;
Matches 125; Conservative 36; Mismatches 40; Indels 0; Gaps 0;
QY 4 GFLVSLGPEGAGKTSVLEALLPILKEGVLTTPREGVGLIGEKIREVILDPSSHQMD 63
Db 3 GILSLGPDGAGTTLVQLPEIQMKREVTPREGGVRAEIRQIILDPKNTIED 62
QY 64 AKTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADLNOF 123
Db 63 SKTEMLFAAARLHMOKMLPALRAGKVIVDFIDSSVAYQYGRDLGVVVDLNYF 122
QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRDREVNRLDLEGLDHLKKVROGYLSLLDKEGNR 183
Db 123 ATDGLKPDLTLYFDITDVALERTMKNRADEVNRLDLERAEMHRKVRGYLEIVAKPEGR 182
QY 184 IVKIDASIPLEQVETTKAVL 204
Db 183 FVKIDASQSLEKVVADTLEVL 203

RESULT 3
Q99WC1 ID Q99WC1 PRELIMINARY; PRT; 205 AA.
AC Q99WC1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TWK PROTEIN
GN TWK OR SA0440.
OS Staphylococcus aureus subsp. aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,

RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003130; BAB41670.1; -.
KW Complete proteome.
SQ SEQUENCE 205 AA; 23424 MW; 71F9C96511FF3A5E CRC64;

Query Match 45.1%; Score 475; DB 2; Length 205;
Best Local Similarity 49.5%; Pred. No. 1.4e-25;
Matches 96; Conservative 40; Mismatches 54; Indels 4; Gaps 3;
QY 7 VSLGEGAGKTSVLEALLPILKEGVLTTPREGVGLIGEKIREVILDPSSHQMDAKT 66
Db 5 ITFEGPEGSGKTVINEVYHRL-VKDYDVIMTREGGVPTGEIRKIVLEGN-DMDIRT 61
QY 67 ELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADLNOFATD 126
Db 62 EAMLFASRRREHLVLPALKEGVLCRDYIDSSLAYQGVARGIGVEEVRLNEFAIN 121
QY 127 GLKPDLTLYFDIEVEEGLARIAANSRDREVNRLDLEGLDHLKKVROGYLSLLDKEGNRIVK 186
Db 122 GLYDPTLYLVNVAEVRERIKNR-RDQNRDLQEDLKFHEKVIIEGYQEIHHNESQRFKS 180
QY 187 IDASIPLEQVETTKAVL 200
Db 181 VNAQPLENVVEDT 194

RESULT 4
Q9X908 ID Q9X908 PRELIMINARY; PRT; 1089 AA.
AC Q9X908;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE INTEGRAL MEMBRANE PROTEIN WITH KINASE ACTIVITY.
GN SCH5.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL035636; CAB38479.1; -.
DR HSSP; P37345; STMP.
DR InterPro; IPR003662; sub.transporter.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.


```

KW Kinase.
SQ SEQUENCE 1089 AA; 115739 MW; F75DD5B2055F409E CRC64;

Query Match 35.4%; Score 373; DB 2; Length 1089;
Best Local Similarity 40.9%; Pred. No. 1.5e-17;
Matches 81; Conservative 35; Mismatches 78; Indels 4; Gaps 1;

QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVVLTTRPGGVLTGKIREVILDPSTQMD 63
   ||:||||:||||: ||| : || ||: ||||| :|||: ||| |
Db 509 GFFIALEGGDAGKSTQAEALAEWIRGHEVLTREPGATPVGKRLRSILLDVSSAGLS 568

QY 64 AKTELLYIASRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQFGRLDTEADWLNQF 123
   : ||| | | : ||| | | : ||| | | : ||| | | : ||| | | : ||| | | :
Db 569 HRAEALLYADRAEHVDTVWRPALERGAVVSDRYIDSSVAYQAGRDLSPTETARINRW 628

QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSREVRNLDLEGLDHLKVKRGYLSLLDKGNR 183
   ||: ||| | | : ||| | | : ||| | | : ||| | | : ||| | | : ||| | | :
Db 629 ATNGLPHLTVLLDVAPAEARETERAPD---RLESEPAEFHARVRSFGFTLAAADPGR 684

QY 184 IVKIDASLPLEQVVETK 201
   : ||| | | :
Db 685 YLVVDAGQEPNAVTVVR 702

RESULT 5
QYEQZH7 PRELIMINARY; PRT; 226 AA.
AC QYEQZH7
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE THYMIDYLATE KINASE.
GN TMK.
OS Xanthomonas albilineans.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=29447;
RN SEQUENCE FROM N.A.
RP MEDLINE=20547559; PubMed=11094291;
RX Huang G., Zhang L., Birch R.G.;
RA "Characterization of the acyl carrier protein gene and the fab gene
RT locus in Xanthomonas albilineans."
RL FEMS Microbiol. Lett. 193:129-136(2000).
DR EMBL: AF294440; AAC42374.1; -.
DR InterPro: IPR00062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
KW Kinase.
SQ SEQUENCE 226 AA; 24882 MW; F88D6E277DB7BECF CRC64;

Query Match 34.78%; Score 365; DB 2; Length 226;
Best Local Similarity 43.1%; Pred. No. 6.3e-18;
Matches 93; Conservative 35; Mismatches 74; Indels 14; Gaps 5;

QY 1 MSKGFL-----VSLEGPEGAGKTSVLEALLPILEKGVVLTTRPGGVLTGKIREVIL 55
   ||| : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MSEAVLRHRRVSLGEGGAKTAINAIRDLQAQHEVLTREPGTPLAERIGLL 60

QY 56 D--PSHTQ-----MDAKTELLYIASRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGF 108
   | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 DMAPSSPETAVERLSAETELLVFAARAHVREVIRPALQRGAVVSDRFTDSSVAYQGE 120

QY 109 GRGLDIEADWLNQFATDGLKPDLTLYFDEVEEGLARIAANSREVRNLDLEGLDHLK 168
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 GRGLDRAMIAADLERRAV--GLQPGTLLDLDVQIGRAR--TSGRDLWPDRIESEQDDFQR 178

QY 169 VRCYLSLLDKGNRIKVIDASLPLEQVWETTKAVL 204
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 VRAGFORATQDPQFRFRVIDASQPQAVAAAL 214

```

```

RESULT 6
QYR2M7 PRELIMINARY; PRT; 165 AA.
AC QYR2M7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN TMK.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN SEQUENCE FROM N.A.
RP STRAIN=21506, 25963, AND 21708;
RX MEDLINE=20040673; PubMed=10570195;
RA Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;
RT "Yersinia pestis, the cause of plague, is a recently emerged clone of
RT pseudotuberculosis."
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL: AJ270425; CAB58190.1; -.
DR EMBL: AJ270423; CAB58188.1; -.
DR EMBL: AJ270424; CAB58189.1; -.
DR HSSP; P37345; STMP.
DR InterPro: IPR00062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Kinase; Transferase.
FT NON_TER 1
FT NON_TER 165
SQ SEQUENCE 165 AA; 18220 MW; F9BBEC3F61F2C8EB CRC64;

Query Match 22.1%; Score 233; DB 2; Length 165;
Best Local Similarity 38.6%; Pred. No. 5.2e-09;
Matches 61; Conservative 30; Mismatches 61; Indels 6; Gaps 5;

QY 22 EALLPILEKGV-EVLTTRPGGVLTGKIREVILDPSTQ-MDAKTELLYIASRQHL 79
   : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 3 DTVVATLRAQGINDIVFTREPGGTPLAEKLDKLIKQGDGDEVLTDRAEVLMLYAARVQ-L 61

QY 80 VEKVL-PALEAGKLVIMDRFIDSSVAYQFGRLDIEADWLNQFATDGLKPDLTLYFDI 138
   ||| : ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 VENVIKPALARGSVVWGRHDLSSQAYOGGGINLMTSLRDIVLGEFPDLYLDL 121

QY 139 EVEGLARIAANSREVRNLDLEGLDHLKVKRQGYLSL 176
   |||| | | : | : | | | : | : | | |
Db 122 PPAIGLARARG--ELDRIEQESLAFFETRERYLEL 157

```

```

RESULT 7
QYR2M9 PRELIMINARY; PRT; 165 AA.
AC QYR2M9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN TMK.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN SEQUENCE FROM N.A.
RP STRAIN=24636, AND 383;
RX MEDLINE=20040673; PubMed=10570195;
RA Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;
RT "Yersinia pestis, the cause of plague, is a recently emerged clone of
RT pseudotuberculosis."
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL: AJ270428; CAB58193.1; -.

```

Qy	139	EVEGLARIANSRENVNRLDLEGLDLHKVKRGYLSL	176
Dd	122	PPVIGLARARG--ELDRIEQESLAFFERTTRYLEYL	157
		: : : : :	
RESULT	9		
ID	Q9RIB4	PRELIMINARY;	PRT; 165 AA.
AC	Q9RIB4;		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)		
DE	THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).		
GN	TMK.		
OS	Yersinia pseudotuberculosis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_TaxId=633;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=32817;		
RC	MEDLINE=20040673; PubMed=10570195;		
RA	Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;		
RT	"Yersinia pestis, the cause of plague, is a recently emerged clone of		
RT	pseudotuberculosis.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).		
DR	EMBL; AJ270422; CAB58213.1; -.		
DR	HSP: P37345; STMP		
DR	InterPro; IPR000062; Thymidylate_kin.		
DR	Fram; PF02223; Thymidylate_kin; 1.		
KW	PROSITE; PS01331; THYMIDYLATE_KINASE; 1.		
KW	Kinase; Transferase.		
FT	NON_TER 1		
FT	NON_TER 165 165		
SQ	SEQUENCE 165 AA; 18103 MW; CD385EB9ACD23419 CRC64;		
		Query Match 21.9%; Score 231; DB 2; Length 165;	
		Best Local Similarity 36.6%; Pred. No. 7.2e-09;	
		Matches 60; Conservative 33; Mismatches 65; Indels 6; Gaps	
Qy	22	EALLPILEEGKV-EVLTTREPGGVLIGEKTREVILDPSTHQ-MDAKTELIIYTASRRQHL	79
Dd	3	TVAVALAAGAGINDIVTFTEPGGTPLAEKLRLDIKOGIDGVLTRAEVLMLYAARVQ-L	61
		: : : : : : : : : : : : : : : :	
Qy	80	VEKVL-PALEAGKLVIMDRIDSVAVGFGRLDIEADWLNAQFATDGLKPDLTLYFDI	138
Dd	62	VENVIKPALARGSVWGDRHDLSSQAVQGGRGIDSOLMASLRDTVLGFEFPDLTYLDL	121
		: : : : : :	
Qy	139	EVEGLARIANSRENVNRLDLEGLDLHKVKRGYLSLDKKEG	182
Dd	122	PPAVGLARARG--ELDRIEQESLAFFERTTRYLEYLATSDAS	163
		: : : : : : :	
RESULT	10		
ID	Q9RIE6	PRELIMINARY;	PRT; 165 AA.
AC	Q9RIE6		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)		
DE	THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).		
GN	TMK.		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_TaxId=632;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=554;		
RC	MEDLINE=20040673; PubMed=10570195;		
RA	Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;		
RT	"Yersinia pestis, the cause of plague, is a recently emerged clone of		

Qy	80	VEKVL-PALBAGKLIVIMDRFSSVAYQGGRGLDIEADWLNQFATDGLKPDLTLYFDI	130
Db	62	VENVIKPALARGSWVGDRHDLSSAQYGGGGRGIDSQLMASTRDTVLGEFRPDLTYLIDL	121
Qy	139	EVEEGLARIANSREVNRLDLEGLDLHLHKVKRGYLISL	176
Db	122	PPAVGLARARG--ELDRIEQESLAFFERTRARYLEL	157
RESULT	12		
Q49001	ID	Q49001 PRELIMINARY; PRT; 98 AA.	
AC	Q49001;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	ATP-BIND. PYRIMIDINE KINASE (FRAGMENT).		
OS	Mycoplasma capricolum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;		
OC	Entomoplasmataceae.		
OX	NCBI_TaxID=2095;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 27343(KID);		
RX	MEDLINE=96059641; PubMed=7476192;		
RA	Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,		
RA	Gilbert W., Gillevet P.M.;		
RT	"Exploring the Mycoplasma capricolum genome: a minimal cell reveals		
RT	its physiology." ;		
RL	Mol. Microbiol. 16:955-967(1995).		
DR	EMBL; Z33079; CAA83746.1; -.		
DR	HSSP; P37345; 5TMP.		
FT	NON_TER 98		
SQ	SEQUENCE 98 AA; 11449 MW; FA597FA1685A0822 CRC64;		
Query Match	19.1%; Score 201; DB 2; Length 98;		
Best Local Similarity	51.2%; Pred. No. 4.1e-07;		
Matches	42; Conservative 15; Mismatches 25; Indels 0; Gaps		
Qy	7	VSLEGPEGAGKTSVLEALLPIEEKGVFLTTREPQGVLIKEKIREVILDPSTHQMDAKT	66
Db	3	ITFGDMDSGSKTTALLKVKEELERNLYKVLTITREPQGVETAEIQIRILDNNKNKDMAWT	62
Qy	67	ELLLYTASRRQHLLVEKVLPALE 88	
Db	63	EALLFTASRNQHLQVKYIKPALE 84	
RESULT	13		
O83373	ID	O83373 PRELIMINARY; PRT; 208 AA.	
AC	O83373;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
GN	THYMIDYLATE KINASE (TMK).		
DN	TP0354.		
OS	Treponema pallidum.		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.		
OX	NCBI_TaxID=160;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NICHOLS;		
RX	MEDLINE=98332770; PubMed=9665876;		
RA	Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,		
RA	Dodson C.M., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,		
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,		
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,		
RA	McDonald L., Artich P., Bowman C., Cotton M.D., Fujii C., Garland S.,		
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,		
RA	Venter J.C.;		
RT	*Complete genome sequence of Treponema pallidum, the syphilis		

[illegible]

<hr/>					
Query Match	16.5%;	Score 174;	DB 2;	Length 208;	
Best Local Similarity	30.0%;	Pred. No. 8.4e-05;			
Matches	61;	Conservative 36;	Mismatches 80;	Indels 26;	Gaps 8;
<hr/>					
QY	7	VSLEGGAGKTSVLEALLPILBEKGVEVLTTREPGVLIGEKIREVIDLPSHTQMDAKT	66		
		: :	: :	:	: :
Db	8	VFEGIDGTSTQLRALERHQAR-KDMVFTQEPTGTEGLTRDVL--QKRVISSKA	64		
		:	:	:	:
QY	67	ELLYLTASRQHLL--VEKVLPALEAGKLYIMDRFIDISSVAYQGFGRLDIEADINQFA	124		
		: :	:	:	:
Db	65	LGLLFADRHELEGAGGINDCLAEGKIVLCRDYVFSSLYIQGM-----AVS--GSFA	115		
		:	:	:	:
QY	125	TDLKL----PDLPFLYDIEVEEGLRTAANSRDENVRLDEGLDLHKVKRQGLSLLDK-	179		
		: :	: :	: :	: :
Db	116	YELNKEFPPLPEVVFYDAPIEVCRITA---RGLQTELYEYTSFOEKARKGYETIFRC	172		
		: :	: :	: :	: :
QY	180	----EGNRIVKIDASLPLOQVE	198		
		:	:	:	:
Db	173	RHLYPAMKVIEDAREIEVHVE	195		
<hr/>					
RESULT	14				
O51733					
ID	O51733	PRELIMINARY;	PRT;	257 AA.	
AC	O51733;				
OT	01-JUN-1998 (TREMBLrel. 06, Created)				
<hr/>					
RESULT	15				
Q9C2R7					
ID	Q9C2R7	PRELIMINARY;	PRT;	309 AA.	
AC	Q9C2R7;				
DT	01-JUN-2001 (TREMBLrel. 17, Created)				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	RELATED TO THYMIDYLATE KINASE.				
GN	104H10.70.				
OS	Neurospora crassa.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Sordariales; Sordariaceae; Neurospora.				
OX	NCBI_Taxid=5141;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Schulte U., Aign V., Hoheisel J., Brandt P., Farmann B., Holland R.,				
RA	Nyakatura G., Mewes H.W., Mannhaupt G.;				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	German Neurospora genome project;				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL; AL513410; CAC28573.1;				
KW	Kinase.				
SQ	SEQUENCE 309 AA; 33757 MW; AC2E55FC6B795DD1 CRC64;				

[illegible]

Search completed: February 15, 2002, 02:01:43
Job time: 386 sec

Query Match 13.8%; Score 145; DB 2; Length 257;
Best Local Similarity 26.7%; Pred. NO. 0.011;
Matches 48; Conservative 35; Mismatches 79; Indels 18; Gaps 8;
QY 9 LEGPEGAGTGVLEALLPILEKGVVEVLTTRPPGGVLGCEKIREVILDPSTOMDAKTEL 68